HAL open science

Tatouage informé de signaux audio numériques

Cléo Baras

To cite this version:

Cléo Baras. Tatouage informé de signaux audio numériques. domain_other. Télécom ParisTech, 2005. English. NNT : . pastel-00001790

HAL Id: pastel-00001790

https://pastel.hal.science/pastel-00001790

Submitted on 19 Jun 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TELECOM PARIS
école nationale supérieure des télécommunications

École Doctorale d'Informatique, Télécommunications et Électronique de Paris

Thèse

présentée pour obtenir le grade de docteur

de l'Ecole Nationale Supérieure des Télécommunications

Spécialité : Signal et Images

Cléo BARAS

Tatouage informé de signaux audio numériques

Soutenue le 7 décembre 2005 devant le jury composé de

Pierre Duhamel	Président
Madeleine Bonnet	Rapporteurs
Gang Feng	
Séverine Baudry	Examinateurs
André Gilloire	
Nicolas Moreau	Directeur de thèse

Résumé

Le tatouage des signaux audio numériques, outre ses applications de protection de la propriété intellectuelle, révèle la possibilité d'utiliser le signal audio numérique comme un support de communication : une information peut être insérée imperceptiblement dans un signal audio et être ainsi diffusée en suivant le même canal de transmission que celui emprunté par le signal audio. Concevoir un système de tatouage performant est l'objectif de ces travaux de thèse. Cette performance relève de contraintes et d'enjeux spécifiques : l'insertion du tatouage doit (1) être transparente (inaudible), (2) offrir un débit de transmission le plus élevé possible pour une fiabilité de transmission la meilleure possible tout en étant robuste à un ensemble de perturbations (traitements licites qui peuvent être effectués sur un signal audio lors de son parcours dans un réseau de diffusion) et (3) être peu coûteux en temps de calcul pour permettre une implémentation temps-réel.

Pour satisfaire au mieux la contrainte de transparence, un module de contrôle d'inaudibilité innovant, adjacent au modèle psychoacoustique utilisé pour la mise en forme spectrale du tatouage, est proposé. Ce module, basé sur l'algorithme de mesure objective de la qualité sonore PEAQ, établit de façon adaptative les conditions de transparence locale du tatouage en fonction du signal audio.

Atteindre une fiabilité de transmission élevée et garantir la robustesse aux perturbations (telles que la compression MPEG, les changements de format, les filtrages, etc.) constituent la majeure partie de ce travail. La prise en compte, ces dernières années, de la connaissance du bruit audio pour la construction du tatouage a permis de très nettes avancées dans ce domaine tant sur la formalisation du problème de tatouage que sur les performances à escompter des systèmes, dit informés, qui en découlent. La majorité des systèmes de l'État de l'Art s'est concentrée sur la recherche d'un tatouage robuste souvent au détriment de la contrainte d'inaudibilité. A l'inverse, nous proposons différentes stratégies d'insertion du tatouage qui établissent en premier lieu l'inaudibilité locale du tatouage et maximisent ensuite la robustesse de sa transmission aux perturbations.

Les performances du système et l'efficacité des stratégies d'insertion proposées sont établies sur des signaux audio réels et comparées à des systèmes de la littérature : une transmission "audible mais non gênante" avec un TEB de $0,001$ peut être obtenue pour un débit de 170 bit/s en l'absence de perturbations (contre 80 bit/s pour une insertion non informée) ; le débit devra être fixé à 75 bit/s pour garantir la robustesse aux perturbations non désynchronisantes et à 125 bit/s pour résister à une opération désynchronisante.

Abstract

Audio data hiding groups technics that aim at embedding a binary information in an audio signal without introducing perceptual degradation. Well known applications of audio data hiding, namely watermarking, are related to copyright protection, privacy and secrecy. In the last decade a new application field is emerging on the fringe of the previous two. It consists in using the audio signal as a transmission channel : some information can be embedded in an imperceptible manner into the audio signal and be broadcasted by following the same broadcast channel than the audio signal. The complete design of an efficient audio data hiding system is the purpose of this thesis. The efficiency of such an audio data hiding system is related to specific constraints and requirements, that are the following : (1) the inaudibility of the embedded signal, (2) the highest transmission rate with the lowest transmission reliability, (3) the robustness to a various set of perturbations (that are all the licit operations that can be applied on an audio signal during its transmission over a broadcast network, such as MPEG compression, format changes or filtering), and (4) a low computational cost for real time applications.

To satisfy the inaudibility constraint, we propose an innovative procedure. It integrates the classical use of a psychoacoustical model to shape the watermark according to the simultaneous masking characteristics of the audio signal. Furthermore, it adjusts the adaptive parameters of the psychoacoustical model with respect to an objective evaluation of the perceptual distorsion between the audio signal and the watermarking signal, computed with the PEAQ algorithm.

The major part of this work deals with reaching the transmission reliability and the robustness to channel perturbations requirements. The so-called informed embedding strategies, proposed by the State-Of-The-Art, have already proved their efficiency for the design of an audio watermarking system. These strategies use the *a priori* knowledge of the audio signal during the embedding process to choose an appropriate watermark signal that conciliates the inaudibility constraint and the robust transmission of the embedded information. Nevertheless, most of the State-Of-The-Art systems focus on the robustness constraint, being careless about the inaudibility constraint. Therefore, we propose several embedding strategies that ensures the local inaudibility constraint of the embedded information before maximizing the robustness of the transmission to channel perturbations.

System performance in terms of inaudibility of the watermark, transmission reliability with respect to various perturbations and computational cost is measured on real audio signals. It is compared to the performance of State-Of-The-Art systems in order to evaluate the efficiency of the proposed informed embedding strategies. Our system ensures a "perceptible, but not annoying" transmission with a $0,001$ BER at 170 bps transmission rate (in comparison to 80 bps for a non-informed system) when the channel is free from perturbation ; the transmission rate has to be decreased to 75 bps to ensure the system robustness to various perturbations (even desynchronizing perturbations).

Remerciements

Si faire le choix d'une thèse m'apparaissait comme une évidence, la mener à bien s'est avéré quelque peu ardu. Résumer ces trois ans de travail pourrait ressembler à une recette de cuisine - un subtil mélange de passion pour la connaissance, de patience et d'obstination, le tout pimenté de nombreuses questions et saupoudré de quelques réponses, agrémenté de doutes et fouetté énergiquement avec Matlab. L'émulsion n'aurait en tout cas jamais prise sans le concours et le soutien de nombreuses personnes, auxquelles j'aimerais, dans ces quelques lignes, exprimer ma reconnaissance.

Mes remerciements s'adressent en premier lieu aux membres du jury pour l'intérêt qu'ils ont porté à mon travail.

Je remercie Pierre Duhamel, directeur de recherche CNRS au LSS [1], pour m'avoir fait l'honneur de présider mon jury de thèse, mais également pour les orientations scientifiques qu'il m'a suggérées au cours des conférences où nous avons eu l'occasion de nous croiser.

Je souhaite vivement remercier Madeleine Bonnet, Professeur à Paris V et Gang Feng, Professeur INPG [2], tant pour avoir accepté d'être rapporteurs de ce manuscrit que pour m'avoir aiguillée et m'avoir soutenue au cours de mon parcours dans l'univers de la recherche.

Enfin, je remercie Séverine Baudry et André Gilloire pour avoir accepté d'examiner de cette thèse.

Je souhaite exprimer ma profonde reconnaissance à Nicolas Moreau, mon directeur de thèse. Au cours de ces quatre années, il a su m'apporter son expérience de la recherche, me prodiguer écoute et soutien tant sur le plan scientifique que pour sur le plan humain en faisant preuve d'une disponibilité sans égale. J'ai toujours été portée par la confiance qu'il m'a offerte et j'espère l'avoir honoré par ce travail.

Je remercie chaleureusement Premyslaw Dymarski, qui, en dépit de la distance, a fortement contribué à ces travaux de thèse à travers de nombreuses discussions estivales ou mailistiques.

Je souhaite également remercier Ghassan Kawas Kaleh, Professeur à l'ENST, sans qui je n'aurais pu ouvrir la voie aux démonstrations théoriques attenantes à cette thèse.

Enfin, je souhaite exprimer ma reconnaissance à Bernard Robinet, directeur de l'EDITE de Paris, sans qui cette aventure n'aurait sans doute pas débutée.

L'accueil et l'ambiance quotidienne du département TSI de l'ENST a fortement contribué au

1. Laboratoire des Signaux et Systèmes
2. Institut National Polytechnique de Grenoble

plaisir que j'ai éprouvé à être thésarde et je remercie toutes celles et tous ceux qui y ont pris part.

Je pense plus particulièrement à Henri Maître et à Yves Grenier, ancien et nouveau directeur du département, qui m'ont permis de m'y séjourner quelque temps.

Je ne saurais oublier Laurence, Catherine et Patricia, les trois fées magiques des thésards, Fabrice et Sophie-Charlotte, faiseurs de miracles informatiques, Gaël, Bertrand, Béatrice, Ioana et Gérard, prophètes du traitement du signal en terre dareaulienne.

Ces remerciements vont aussi à Sonia Larbi, avec qui j'ai eu le plaisir de partager, au cours de ses trop brefs séjours parisiens, de nombreuses discussions fructueuses *entre tatoueuses passionnées*.

Je souhaite adresser enfin une pensée toute particulière à l'équipe des thésards de TSI (bien que certains aient déjà soutenus depuis quelques temps) et à tous ceux qui les "supportent" (le jour comme la nuit). J'espère que tous ces bons moments passés au labo (et ailleurs !) auront de nombreuses occasions de se reproduire. Merci donc à Slim (qui a si patiemment répondu à toutes mes questions "idiotes" et a, je l'espère, partagé plus qu'un bureau) ; à Greg (co-bureau des débuts pour partager mes rêves de province...) ; à Chloé (ma clarinettiste-saxophoniste préférée), les Cloportes Thomas, Julie, Mathieu et leurs nombreux managers Raphaël, Remi, et Edouardo ; à Christophe (pour nos pauses-clope tardives mais si instructives) ; à Miguel, Loïs (avec tout le courage qu'on leur souhaite en cette phase difficile de rédaction) ; à Valentin, Pierre et Simone (avec ce même courage pour atteindre la rédaction) ; à Roland et Thomas ; à Téo, Maria et Nancy (dignes représentantes de la gente - trop rare - des doctorantes) ; à tous les stagiaires ; à Caro, Seb, Lionel et Tize (mes quatre compères de bateau...).

Il est difficile de décrire en quelques mots le soutien que ma famille et mes amis ont su m'apporter ; je préfère simplement leur dire un immense merci. Merci à Papoune, à Maman, à ma sister Diane, à Aude & Guigui, Caro & Antoine, Barbara, P.J & Isma, Julien & Julia, Marie-Paule & Denis, Francesco, Ben, Mariane, Hervé, Fred, Emilie, Virginie & Anthony, Benoît, Ewan, Ingrid & François, aux bretons et surtout, à mon breton préféré Eric.

Table des matières

Résumé **i**

Abstract **ii**

Remerciements **iv**

Table des matières **vii**

Table des figures **xi**

Liste des tableaux **xv**

Liste des abréviations **xvii**

Liste des notations **xix**

Introduction générale **1**

I Enjeux et contexte d'étude **5**

1 Le signal audio, un nouveau support de communication **7**
1.1 Tatouage et données cachées 7
1.2 Applications 8
1.3 Objectifs et contraintes 9

2 Critères et mesures de performances d'un système de tatouage **11**
2.1 Mesurer la transparence du tatouage 11
2.2 Évaluer la fiabilité et la robustesse de la transmission 14
2.3 Mesurer le coût en temps de calcul 16
2.4 Protocole expérimental 17

3 Etat de l'Art **21**
3.1 Modèle général d'un système de tatouage 22
3.2 Spécificités des systèmes non informés 29
3.3 Spécificités des systèmes de tatouage informés 32
3.4 Spécificités des systèmes pour la synchronisation 43

Conclusion **49**

II Un système de tatouage audio **51**

Introduction **53**

4 Éléments de théorie des communications numériques **55**
4.1 Message numérique 55
4.2 Modulation 56
4.3 Récepteur optimal 57
4.4 Performances théoriques 63

5 Construction d'un système de tatouage audio **69**
5.1 Modulation 70
5.2 Réception 80

6 Paramétrisation du système de tatouage **87**
6.1 Filtres de la chaîne de tatouage 87
6.2 Réglage de la bande passante : impact de la compression MPEG 91
6.3 Seuil de masquage 94
6.4 Contrôle de l'inaudibilité par un facteur d'amplitude α adapté au signal audio . . 96

7 Performances du système de tatouage **103**
7.1 Conditions de transmission 103
7.2 Performances expérimentales 107

Conclusion **115**

III Vers un système de tatouage informé **117**

Introduction **119**

8 Conditions d'une transmission inaudible et robuste **121**
8.1 Copie locale du récepteur à l'émetteur 121
8.2 Conditions d'une transmission inaudible et robuste 124
8.3 Modèle de la stratégie d'insertion informée 127

9 Tatouage M-aire de robustesse maximale aux perturbations **129**
9.1 De la nécessité d'une stratégie informée 129
9.2 Vers un dictionnaire structuré 132
9.3 Tatouage de robustesse maximale aux perturbations additives 144
9.4 Système de tatouage en boucle fermée avec dictionnaire structuré et choix d'un tatouage de robustesse maximale aux perturbations 166

10 Tatouage CDMA informé **173**
10.1 De la nécessité d'une stratégie d'insertion CDMA informée 173
10.2 Stratégie d'insertion CDMA informée 176
10.3 Application au système de tatouage en boucle fermée 180

Conclusion **187**

IV Synchronisation **189**

11 Mécanisme de synchronisation pour des opérations désynchronisantes à forte dérive **191**

11.1 Modélisation de la désynchronisation 192
11.2 Mécanisme de synchronisation 194
11.3 Performances 205

Conclusions et perspectives **208**

Bibliographie **212**

Publications **220**

Table des figures

3.1 Schéma de principe d'un système de tatouage 22
3.2 Modèle d'un système de tatouage 28
3.3 Système de tatouage non informé de Larbi 30
3.4 Système de tatouage non informé de Cvejic *et al.* 30
3.5 Performances des systèmes de tatouage de Larbi et de Cvejic *et al.* 32
3.6 Interprétation géométrique de la stratégie d'insertion proposée par Miller *et al.* . 35
3.7 Modulation en treillis proposée par Ungerboëck 39
3.8 Interprétation géométrique de la QIM 41
3.9 Performances du système de tatouage informé de Cvejic et Seppänen 43
3.10 Schéma simplifié du système de tatouage pour la mise en évidence des mécanismes de synchronisation 45

4.1 Modèle synthétique de la chaîne de tatouage 56
4.2 Représentation des constellations possibles d'un dictionnaire de 4 vecteurs 62
4.3 Probabilité d'erreur binaire des modulations M-aire pour différentes tailles M d'un dictionnaire orthogonal 65
4.4 Probabilité d'erreur binaire des modulations M-aire pour différentes tailles M d'un dictionnaire biorthogonal 65
4.5 Probabilité d'erreur binaire des modulations CDMA 67

5.1 Vue générale de l'oreille 70
5.2 Seuil d'inaudibilité du son masqué en fonction de l'ordre et du temps d'apparition du son masqué et du son masquant 73
5.3 Fonctions d'étalement pour des bruits à bande étroite 74
5.4 Allure du seuil de masquage obtenu avec trois modèles psychoacoustiques 75
5.5 Système de tatouage lorsque l'émetteur réalise la mise en forme perceptuelle du signal modulé en suivant un MPA 77
5.6 Introduction du récepteur sous-optimal dans le système de tatouage 82
5.7 Schéma du système de tatouage lorsque le récepteur sous-optimal est introduit . 85
5.8 Système de tatouage audio pour une détection aveugle de l'information 86

6.1 Allure des réponses en fréquence du filtre de mise en forme $H(f)$ et du filtre blanchissant $G(f)$ sur une trame de signal de violon 89
6.2 Variation du TEB moyen en fonction de l'ordre P_G du filtre blanchissant 89
6.3 Impact de la compression MPEG sur la DSP des signaux audio et sur les seuils de masquage 92
6.4 Schéma du système de tatouage avec introduction du filtrage passe-bas 93

6.5 Variation de l'ODG moyenne et du rapport masque à signal en fonction du facteur d'amplitude α . . . 97
6.6 ODG en fonction de α . . . 97
6.7 Système de tatouage audio avec module du contrôle de l'inaudibilité . . . 98
6.8 Évolution du gain α et de l'ODG en fonction du temps pour deux signaux . . . 102

7.1 Critères de puissance et d'orthogonalité du dictionnaire de réception en fonction du temps symbole N_s . . . 105
7.2 Évaluation des rapports signaux à bruit dans la chaîne de tatouage . . . 107
7.3 TEBs du système de tatouage non informé de référence en fonction du débit de transmission dans le cas d'un canal sans perturbation . . . 110
7.4 Évolution du TEB à 206 bit/s en fonction des perturbations . . . 112
7.5 Évolution du TEB en fonction du débit face à un canal par filtrage passe-haut . 113
7.6 Variations du coût de réception en fonction du nombre de vecteurs du dictionnaire 114
7.7 Schéma du système de tatouage . . . 116

8.1 Schéma synthétique du système de tatouage avec les deux récepteurs par filtre blanchissant et égalisation . . . 122
8.2 Copie locale du récepteur à l'émetteur . . . 123
8.3 Modèle de la stratégie d'insertion informée pour la maximisation de la robustesse sous contrainte d'inaudibilité . . . 127

9.1 Mise en évidence des régions de distorsion acceptable et de détection . . . 131
9.2 Schéma du système de tatouage . . . 133
9.3 Configurations des constellations engendrées par le dictionnaire structuré . . . 135
9.4 Comparaison des TEBs obtenus par simulation et des tracés théoriques de la probabilité d'erreur . . . 140
9.5 Probabilité d'erreur binaire de la chaîne informée selon le modèle de dictionnaire de Costa en fonction du débit . . . 142
9.6 Comparaison des performances du système non informé et de la stratégie d'insertion par dictionnaire structuré . . . 145
9.7 Schéma du système de tatouage . . . 146
9.8 Mise en évidence du signal modulé de robustesse maximale aux perturbations dans le cas d'un dictionnaire de deux vecteurs orthogonaux . . . 149
9.9 Interprétation géométrique de l'algorithme d'Uzawa itéré . . . 152
9.10 TEBs en fonction de N_bRSB obtenus par le modèle non informé et les 3 stratégies d'insertion informées (l'algorithme Petit Pas, l'algorithme d'Uzawa itéré et l'algorithme probabiliste) ; différences entre les TEBs obtenus par l'algorithme d'Uzawa et les TEBs des algorithmes Petit Pas et probabiliste en fonction du paramètre N_bRSB . . . 159
9.11 Critères de puissance, de robustesse et de RSB pour les trois stratégies d'insertion informée et la stratégie non informée . . . 160
9.12 TEBs en fonction du paramètre N_bRSB et de la taille M du dictionnaire pour le système non informé et le système informé basé sur l'algorithme Petit Pas . . . 161
9.13 Comparaison des TEBs du système avec stratégie d'insertion informée basée sur l'algorithme Petit Pas aux TEBs du système non informé et de la stratégie d'insertion par dictionnaire structuré . . . 165
9.14 Comparaison des TEBs du système informé final avec les systèmes non informés et informés . . . 168

9.15 TEBs à 206 bits/s du système non informé et informé en fonction des perturbations de canal 169
9.16 TEBs en fonction du débit lorsque le canal est dégradé par un filtre passe-haut . 170

10.1 Mise en évidence des régions de distorsion acceptable et de détection 176
10.2 TEBs en fonction du paramètre N_bRSB dans le cas du système non informé et dans le cas de la stratégie d'insertion informée avec modulation CDMA 179
10.3 Comparaison de la stratégie d'insertion informée et de la stratégie non informée pour les modulations CDMA à travers les critère de puissance, de robustesse minimale et de robustesse maximale 179
10.4 Comparaison des performances du système de tatouage CDMA non informé et de la stratégie d'insertion informée 183
10.5 TEBs du système CDMA informé et du système non informé à un débit de 206 bit/s en fonction des différentes perturbations de canal 184
10.6 TEBs en fonction du débit lorsque le canal est dégradé par un filtre passe-haut . 185

11.1 Schéma de principe du mécanisme de synchronisation pour les phases d'acquisition et de poursuite 195
11.2 Structure du signal modulé pour le mécanisme de synchronisation à l'émission et à la réception 196
11.3 Fonctions de corrélation entre le signal reçu $\hat{v}(n)$ et la pattern de synchronisation au voisinage de la localisation de la pattern 198
11.4 Exemple de fonction d'autocovariance $f(n)$ pour la détection des patterns de synchronisation initiale 200
11.5 Erreur d'estimation moyenne de la fréquence d'échantillonnage de réception $\hat{F}_e$ en fonction de la dérive et du degré de traitement de la phase d'initialisation . . 200
11.6 Évolution de l'erreur d'estimation sur la fréquence d'échantillonnage à la réception $\hat{F}_e$ pour deux méthodes de suivi 203
11.7 Mise en évidence des performances du mécanisme de synchronisation 206

Liste des tableaux

2.1 Échelle de dégradation à cinq notes et valeurs de la SDG associée 12
2.2 Signaux audio du corpus 17
2.3 Perturbations prises en compte pour évaluer la robustesse du système de tatouage 18

6.1 Listes des paramètres de la chaîne 88
6.2 Influence de l'erreur d'estimation des filtres de la chaîne de tatouage 90
6.3 Estimation de l'erreur d'estimation des filtres de la chaîne de tatouage par la mesure de la déviation quadratique 90
6.4 Influence de la largeur de bande F_c sur les TEBs de la chaîne de tatouage 94
6.5 Mesure de l'erreur d'estimation DQ(H) du filtre de mise en forme en fonction de la largeur de bande F_c 94
6.6 Performances de trois modèles psychoacoustiques en terme de TEB, de robustesse et de temps de calcul 96
6.7 Paramétrisation du système de tatouage 101

7.1 Valeurs moyennes et écart-type de l'ODG en fonction de la modulation et de l'utilisation du module de contrôle de l'inaudibilité 108
7.2 Liste des perturbations et de leur numéro associé 111
7.3 Coût en temps de calcul du module de contrôle d'inaudibilité à l'émetteur 113

9.1 Valeurs moyennes de l'ODG et écart-type pour le système de tatouage non informé de référence et les deux systèmes en boucle fermée 144
9.2 Coût en temps de calcul des trois stratégies d'insertion informée 161
9.3 Valeurs moyennes de l'ODG et écart-type pour le système de tatouage non informé de référence et les deux systèmes en boucle fermée 165
9.4 Valeurs moyennes et écart-type de l'ODG pour deux récepteurs et quatre systèmes de tatouage informés et non 168
9.5 Liste des perturbations et de leur numéro associé 169
9.6 Coût en temps de calcul des différentes configurations d'émission et de réception 170

10.1 Valeurs moyennes et écart-type de l'ODG pour le système de tatouage avec modulation CDMA non informé et informé 183
10.2 Liste des perturbations et de leur numéro associé 184
10.3 Coût en temps de calcul des différentes configurations d'émission et de réception 185

11.1 Coût en temps de calcul des deux récepteurs lorsque le mécanisme de synchronisation est ou non activé 207

Listes des abréviations

AAC	Audio Advanced Coding
AQUAFLUX	AQUA marquage de FLUX multimédia (audio et vidéo) pour leur transmission, diffusion, circulation, distribution sur les réseaux hétérogènes de télécommunications
AR	Auto-Régressif
ARTUS	Animation Réaliste par Tatouage à Usage des Sourds
BBAG	Bruit Blanc Additif Gaussien
BCH	Bose-Chaudhuri-Hochquenghem
bit/s	bits par seconde
BPSK	Binary Phase Shift Keying
CA	Critère d'amplitude
CAN	Conversion Analogique Numérique
CDMA	Code Division Multiple Access
CCE	Codes Correcteurs d'Erreur
CEC	Concatenated Extended Complementary
CELP	Code Excited Linear Prediction
CNA	Conversion Numérique Analogique
CO	Conservation d'Orthogonalité
CP	Conservation de Puissance
DC-QIM	Distorsion Compensation Quantization Index Modulation
DCT	Discrete Cosinus Transform
DHT	Discrete Hartley Transform
DIX	Distorsion IndeX
DLL	Delay-Locked Loop
DQ	Déviation Quadratique
DSP	Densité Spectrale de Puissance
DWT	Discrete Wavelet Transform
ENIT	École Nationale d'Ingénieurs de Tunis
ENST	École Nationale Supérieure de Télécommunications
EQM	Erreur Quadratique Moyenne
IES	Interférence Entre Symboles
iid	Indépendant et Identiquement Distribué
IVDS	Interactive Video and Data Services
ISS	Improved Spread-Spectrum

JANIS	Just Another N-order side-Informed Scheme
LPC	Linear Prediction Coding
MCLT	Modulated Complex Lapped Transform
MDCT	Modulated Discret Cosinus Transform
MDL	Minimum Description Length
MIPS	Million Instructions Per Second
MPA	Modèle PsychoAcoustique
MPEG	Motion Picture Expert Group
NMR	Noise to Mask Ratio
OASE	Objective Audio Signal Evaluation
ODG	Objective Difference Grade
PAQM	Perceptual Audio Quality Measure
PB	Passe-bas
PEAQ	Perceptual Evaluation of Audio Quality
PERCEVAL	PERCeptual EVALuation of the quality of audio signals
PL	Prédiction Linéaire
PNSR	Peak Signal to Noise Ratio
POM	Perceptual Objective Measurement
QIM	Quantization Index Modulation
QMF	Quadrature Mirror Filter
RMS	Rapport de puissance entre le seuil de Masquage et le Signal audio
RNRT	Réseau National de Recherche en Télécommunications
RSB	Rapport Signal à Bruit
SCS	Scalar Costa Scheme
SDG	Subjective Difference Grade
SDMI	Secure Digital Music Initiative
SPL	Sound Pression Level
SQP	Sequential Quadratique Programming
SNR	Signal to Noise Ratio
STDM	Spread Transform Dither Modulation
TDL	Tau-Dither Loop
TEB	Taux d'Erreur Binaire
TFD	Transformée de Fourier Discrète
TNT	Télévision Numérique Terrestre
TSI	Traitements des Signaux et des Images
UFR	Unité de Formation et de Recherche
UMTS	Universal Mobile Telecommunication System
UIT	Union Internationale des Télécommunications

Notations

Nomenclature

$x(n)$	Signal audio original
$y(n)$	Signal audio tatoué
$t(n)$	Signal de tatouage
$v(n)$	Signal modulé
$u(n)$	Signal modulé mis en forme spectralement
$p(n)$	Bruit introduit par le canal de transmission
$\hat{y}(n)$	Signal audio tatoué détérioré
$r(n)$	Signal audio blanchi
$\hat{v}(n)$	Signal audio tatoué détérioré puis blanchi
$\mathbf{c}$	Vecteur de corrélation
$\tilde{x}(n)$	Composante audio à l'entrée du corrélateur
$\tilde{v}(n)$	Composante du signal modulé à l'entrée du corrélateur
$\tilde{p}(n)$	Composante du bruit de canal à l'entrée du corrélateur
$\mathcal{D}$	Dictionnaire d'émission
$\hat{\mathcal{D}}$	Dictionnaire de réception
$d(n)$	Forme d'onde du dictionnaire d'émission
$\hat{d}(n)$	Forme d'onde du dictionnaire de réception
$p_i(n)$	Pattern de synchronisation pour la phase d'initialisation
$p_a(n)$	Pattern de synchronisation pour la phase d'acquisition
R	Débit de transmission
N_s	Temps symbole
N_b	Temps bit
N_i	Durée de la pattern de synchronisation $p_i(n)$
N_a	Durée de la pattern de synchronisation $p_a(n)$
N_m	Durée d'un message
b	Information binaire à tatouer
B	Nombre de bits total
s	Information à tatouer sous forme symbolique

$\mathbf{b}$	K-uplet binaire
L	Nombre de K-uplet binaire
L_m	Nombre de symbole par message
K	Nombre de bits encodé par symbole
Q	Nombre de vecteurs du dictionnaire représentant un même symbole
M	Nombre de vecteurs du dictionnaire
$\hat{s}$	Symbole reçu
$\hat{\mathbf{b}}$	K-uplet binaire reçu
V	Variable aléatoire modélisant l'état du signal modulé $v(n)$
X	Variable aléatoire modélisant le signal audio $x(n)$
$\mathcal{N}(a, b)$	Loi normale de moyenne a et de déviation standard b
P_e	Probabilité d'erreur binaire
P_c	Probabilité d'une décision correcte
P_s	Probabilité d'erreur par symbole
$G(f)$	Filtre blanchissant associé au signal audio original x
$\hat{G}(f)$	Filtre blanchissant associé au signal audio tatoué détérioré $\hat{y}$
$H(f)$	Filtre de mise en forme spectrale associé au signal audio original x
$\hat{H}(f)$	Filtre de mise en forme spectrale associé au signal audio tatoué détérioré $\hat{y}$
$W(f)$	Filtre de Wiener
$\mathbf{g}$	Coefficients du filtre $G(f)$ sous forme vectorielle
$g(n)$	Réponse impulsionnelle du filtre $G(f)$
$h(n)$	Réponse impulsionnelle du filtre $H(f)$
$\mathbf{w}$	Coefficients du filtre $W(f)$ sous forme vectorielle
P_H	Ordre des filtres de mise en forme spectrale
P_G	Ordre des filtres blanchissants
P_W	Nombre de coefficients de la partie causale de $W(f)$
Q_W	Nombre de coefficients de la partie non-causale de $W(f)$
N_H	Taille des fenêtres d'analyse pour le calcul du modèle psychoacoustique
N_G	Taille des fenêtres d'analyse pour le calcul du filtre blanchissant
$S_M(f)$	Seuil de masquage
$S_a(f)$	Seuil d'audition absolue
α	Facteur d'amplitude

Formalisme mathématique

$x(n)$	Signal x
$\mathbf{x}$	Vecteur de dimension N_s extrait du signal $x(n)$
$\mathbf{x}[i]$	i-ième composante du vecteur $\mathbf{x}$
X	Variable aléatoire X
$\lvert x \rvert$	Valeur absolue de x
$\text{sign}(x)$	Signe de x
$[x]$	Arrondi de x
$\lfloor x \rfloor$	Partie inférieure de x : arrondi à la valeur inférieure
$\lceil x \rceil$	Partie supérieure de x : arrondi à la valeur supérieure
$E[X]$	Espérance de la variable aléatoire X
$\text{Prob}(a)$	Probabilité de l'événement a
$\text{Prob}(a\|b)$	Probabilité de l'événement a conditionnellement à l'événement b
$\mathbf{x}^t\mathbf{y}$	Corrélation (produit scalaire) entre $\mathbf{x}$ et $\mathbf{y}$: $\mathbf{x}^t\mathbf{y} = \sum_{n=0}^{N-1} \mathbf{x}[n]\mathbf{y}[n]$
$\mathbf{x} \otimes \mathbf{y}$	Produit composantes à composantes des vecteurs $\mathbf{x}$ et $\mathbf{y}$: si $\mathbf{z} = \mathbf{x} \otimes \mathbf{y}$, $\forall i, \mathbf{z}[i] = \mathbf{x}[i]\mathbf{y}[i]$
$\lvert\lvert\mathbf{x}\rvert\rvert$	Norme du vecteur $\mathbf{x}$: $\lvert\lvert\mathbf{x}\rvert\rvert^2 = \sum_{n=0}^{N-1} x^2$
$x(n) \star y(n)$	Produit de convolution entre $x(n)$ et $y(n)$: $x(n) \star y(n) = \sum_{p=-\infty}^{\infty} x(n)y(n-p)$
$\sigma_x^2 = \frac{1}{N}\sum_{n=0}^{N-1} x(n)^2$	Puissance du signal $x(n)$
$S_x(f)$	Densité spectrale de puissance d'un processus aléatoire $X(n)$
R_x	Fonction d'autocovariance de $x(n)$
Γ_X	Matrice d'autocorrélation de la matrice X
$\delta(i,j)$	Symbole de Kronecker égal à 1 si $i = j$ et 0 sinon
$k! = \prod_{i=1}^{k} i$	Factorielle de k
$\text{erfc}(x)$	Fonction d'erreur complémentaire définie par $\frac{1}{2}\text{erfc}\left(\frac{x}{\sqrt{2}}\right) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp\left(-\frac{\theta^2}{2}\right) d\theta$

Introduction générale

Notre système auditif néglige certaines parties de l'information contenue dans un signal audio. Cette propriété est utilisée pour tatouer des signaux numériques, c'est à dire pour y inclure une information supplémentaire non perceptible. Cette technique s'est beaucoup répandue, au cours des 10 dernières années, face à la forte demande industrielle en matière de protection de la propriété intellectuelle. Mais plus généralement elle permet d'ajouter dans la transmission des données audio un canal supplémentaire caché. Ce canal sert différents intérêts : il peut véhiculer un label pour aider l'indexation des signaux, des meta-données décrivant le contenu du signal ou des informations à destination d'une application cible. Cette application cible peut être, à l'exemple du projet RNRT [3] ARTUS [4] [Rés03] auquel nous avons participé au cours de cette thèse, l'animation d'un clone virtuel. Ce clone se substitue au sous-titrage destiné aux malentendants, pour le traduire dans une langue des signes (le langage parlé complété). Le tatouage permet alors la transmission des paramètres. En effet, la transmission des paramètres s'effectue par tatouage de la séquence audiovisuelle ; elle peut ainsi suivre le même chemin de diffusion que la séquence audiovisuelle sans traitement spécifique et être transmise au terminal cible (par exemple un ordinateur équipé d'une carte vidéo) en synchronie avec l'émission télévisée.

Dans ce contexte applicatif de transmission de données, le tatouage doit satisfaire à un ensemble de contraintes spécifiques ; obtenir un débit de transmission le plus élevé possible tout en maintenant l'inaudibilité de l'information cachée et la fiabilité de la transmission la meilleure possible est le défit à relever. Le tatouage obtenu se doit également d'être robuste à un ensemble de distorsions (telles que les conversions de format proposées par compression MPEG [5] ou la transmission analogique, les changements d'échelles, etc...) et doit posséder un coût de calcul permettant l'application temps-réel. Ces enjeux spécifiques établissent explicitement la distinction entre tatouage pour la protection de la propriété intellectuelle et tatouage pour la transmission de données auquel nous nous intéressons. Ils définissent également les critères de performances d'un système de tatouage.

L'état actuel de la recherche propose de nombreux systèmes de tatouage, construit sur l'analogie entre chaîne de tatouage et chaîne de communication numérique. Le tatouage est le signal utile porteur de l'information à émettre tandis que le signal audio est le bruit de canal. Les difficultés principales de la construction de ces systèmes résident dans un contexte de transmission peu favorable : un rapport signal à bruit très négatif, une forte instationnarité du canal audio et des perturbations sévères. La prise en compte, ces dernières années, de la connaissance du bruit audio pour la construction du tatouage a permis de très nettes avancées dans le domaine tant sur la formalisation du problème de tatouage que sur les performances à escompter des systèmes,

3. Réseau National de Recherche en Télécommunications
4. Animation Réaliste par Tatouage à Usage des Sourds
5. Motion Picture Expert Group

dit ***informés***, qui en découlent. Concevoir un système de tatouage tient alors d'une stratégie : celle de choisir un tatouage conciliant inaudibilité et transmission robuste de l'information aux perturbations au regard du signal audio. La majorité des systèmes s'est concentrée sur la recherche d'un tatouage robuste souvent au détriment de la contrainte d'inaudibilité. Proposer des stratégies d'insertion qui établissent en premier lieu l'inaudibilité du tatouage et maximisent ensuite la robustesse de sa transmission aux perturbations est toujours d'actualité et constitue l'objectif de cette thèse.

La première partie de ce manuscrit s'attache à définir avec précision les enjeux et les attentes d'un système de tatouage audio dédié à la transmission de données. Les applications et les contraintes de ce système sont décrites au chapitre 1. Elles sont suivies, au chapitre 2, de la définition d'un protocole expérimental qui établit les critères de mesures spécifiques permettant de caractériser les performances d'un système de tatouage dans le contexte de transmission de données. Ces critères feront notamment appel à deux outils mis à disposition de la communauté scientifique, l'algorithme PEAQ [6] [Uni01] pour l'appréciation de la transparence du tatouage et le Stirmark Audio [Lan] pour la robustesse du tatouage aux perturbations. L'étude de l'existant fait l'objet du chapitre 3. Nous présentons dans un premier temps une description synthétique des systèmes de tatouage audio développés dans la littérature. Cette description est suivie d'un état des spécificités des systèmes non informés. Une analyse plus approfondie des stratégies d'insertion informées est ensuite exposée, avec l'objectif de préciser les enjeux et les apports des techniques établies dans la littérature et de cerner leurs faiblesses. Une dernière section est consacrée au problème de la synchronisation dans les systèmes de tatouage. La robustesse de la transmission à cette perturbation très spécifique est obtenue par la mise en place d'outils adaptés, dont nous synthétisons les principes.

La seconde partie de ce manuscrit s'attache à un système de tatouage additif non informé, né de la collaboration entre le département TSI [7] de l'ENST [8] et l'UFR [9] de Mathématiques et d'Informatique de l'Université Paris V pour la thèse de doctorat de L. de C. T. Gomes [Gom02]. Les différentes étapes de la conception de ce système sont reprises voire modifiées avec un important travail de formalisation et de justification. Cette conception est fondée sur un ensemble de pré-requis de la théorie des communications numériques, dont une vue synthétique est présentée dans le chapitre 4. Ces outils, associés aux éléments de la psychoacoustique qui définissent les conditions d'inaudibilité du tatouage, sont ensuite adaptés au contexte particulier de la transmission cachée dans un signal audio ; ils établissent et justifient pas à pas les différentes étapes de la chaîne de tatouage, construite dans le chapitre 5. Cette chaîne construit le signal de tatouage en exploitant notamment une modulation par étalement de spectre (soit de type M-aire soit de type CDMA [10]) et une mise en forme spectrale du tatouage. Cette dernière est dépendante d'un modèle psychoacoustique et d'un module de contrôle de gain adaptatif, au principe novateur, déterminant les paramètres d'inaudibilité spécifiques au signal audio à tatouer. La procédure de détection, initialement basée sur une structure d'égalisation, est également modifiée pour s'approcher d'un récepteur optimal. Elle est définie à un ensemble de paramètres près qu'il convient de choisir avant une mise en œuvre effective sur des signaux réels. Cette paramétrisation fait l'objet du chapitre 6. Pour finir, les performances du système de tatouage non informé sur des signaux réels sont établies au chapitre 7 de sorte à servir de référence pour comparer les stratégies

6. Perceptual Evaluation of Audio Quality
7. Traitements des Signaux et des Images
8. École Nationale Supérieure de Télécommunications
9. Unité de Formation et de Recherche
10. Code Division Multiple Access

informées présentées par la suite.

La troisième partie de ce manuscrit regroupe les contributions majeures de ce travail. Elle présente les différentes stratégies d'insertion informées, c'est à dire les techniques de construction du tatouage dans lesquelles la connaissance *a priori* du signal audio à l'émetteur est prise en compte. Cette connaissance *a priori* du bruit audio est exploitée en introduisant, dans le schéma du système non informé établi dans la seconde partie, une copie locale du récepteur à l'émetteur ; elle donne lieu à un schéma de tatouage en boucle fermée qui permet d'établir, chapitre 8, les conditions d'une transmission inaudible et correcte de l'information face à un bruit modélisant la perturbation et de rétroagir sur le choix du signal de tatouage. Cette rétroaction fait l'objet des chapitres 9 et 10. Le chapitre 9 est dédié aux stratégies d'insertion informées adaptées aux modulations M-aire : le tatouage y est choisi de sorte à maximiser la robustesse de la transmission sous contrainte d'inaudibilité. Deux stratégies sont proposées, chacune proposant une formalisation différente de la robustesse aux perturbations. Les performances de ces stratégies sont étudiées sur un modèle théorique de la chaîne de tatouage puis appliquées à des signaux réels ; elles sont comparées aux performances du système non informé. Le chapitre 10 développe une stratégie d'insertion informée dans le contexte des modulations CDMA. L'objectif est ici encore de choisir un tatouage inaudible et de robustesse maximale aux perturbations en prenant cette fois en compte les contraintes spécifiques de la transmission CDMA. Les performances de cette stratégie sont évaluées sur un modèle de la chaîne de tatouage et sur des signaux réels.

La quatrième et dernière partie de ce manuscrit s'intéresse au cas spécifique des perturbations de canal désynchronisantes. Ce type de perturbations est notamment introduit pour les besoins de la radiodiffusion des signaux audio : la durée temporelle du signal audio tatoué peut être artificiellement diminuée pour respecter des temps de diffusion impartis modifiant ainsi la localisation des informations dans le signal. Le récepteur doit alors mettre en place des outils spécifiques pour permettre au récepteur de se synchroniser sur chaque information. Ces outils peuvent être dissociés de la stratégie d'insertion et donc faire l'objet d'une étude indépendante. Un mécanisme de synchronisation adapté aux opérations désynchronisantes à forte dérive (c'est à dire à forte dilatation ou contraction de l'échelle des temps) et quasi temps-réel est proposé. Ses performances sur le système de tatouage informé sont finalement évaluées sur des signaux réels pour juger de son efficacité.

Première partie

Enjeux et contexte d'étude

Chapitre 1

Le signal audio, un nouveau support de communication

Faciliter l'accès à l'information, développer et créer de nouveaux moyens de communication, promouvoir l'interactivité, sécuriser les échanges - tels sont les enjeux industriels et commerciaux de ces dernières années et de celles à venir. Les techniques d'enfouissement de données, favorisées par l'avènement des documents sous format numérique, offrent des possibilités innovantes dans ce domaine : un document hôte (en particulier un signal audio) peut cacher une information numérique aux applications diverses et servir de nouveau support de transmission. La technique d'enfouissement est alors appelée tatouage.

L'information cachée sert différents intérêts, en fonction de l'application à laquelle elle est destinée. Les fortes demandes industrielles en matière de protection de la propriété intellectuelle la conduisent souvent à devenir une signature ineffaçable même face à un utilisateur malveillant, appelé ***pirate***, qui mettrait tout en œuvre pour la supprimer. En parallèle de ce contexte sécuritaire particulier, cette information permet également d'augmenter le contenu du signal audio, de lui apporter une valeur ajoutée. Ce contexte, souvent occulté par les attentes en matière de protection, fait l'objet de cette thèse.

Ce court chapitre se propose de synthétiser les différentes applications gravitant autour du tatouage pour l'augmentation de contenu et de spécifier les contraintes à satisfaire par un système de tatouage au regard de l'application envisagée.

1.1 Tatouage et données cachées

L'enfouissement de données est l'art d'insérer dans un document, qu'il soit analogique ou numérique, textuel, sonore ou vidéo, un message perceptuellement indétectable. Dans le cas des signaux audio, auxquels s'attache cette thèse, les imperfections du système auditif humain sont mises à profit pour ajouter au son un message inaudible. Le signal audio devient alors un média de communication véhiculant une information cachée.

Les intérêts servis par cette information cachée conduisent à distinguer trois catégories parmi les techniques d'enfouissement de données :

- La ***stéganographie***, en plus d'établir une communication cachée entre un émetteur et un receveur via un signal audio, cherche à cacher l'existence de la communication à une tierce

personne.

- Le ***tatouage*** fait souvent référence à l'insertion d'une information dans un contexte sécuritaire, notamment de protection contre le piratage : l'information souvent associée à une signature ou une marque est liée au contenu même du document. Elle régit la propriété, l'authenticité et les droits d'utilisation du document [CWL01, KM01]. Contrairement à la stéganographie, elle suppose que la communication est connue d'une tierce personne, le pirate. Le tatouage vise justement à adapter la signature en fonction de l'action de ce pirate sur le document : la signature doit être ineffaçable si l'on souhaite lutter contre le piratage des documents ; à l'inverse, elle doit disparaître dès que le pirate modifie le document si l'on souhaite garantir l'intégrité du document.

- Le ***tatouage pour la transmission de données*** permet de véhiculer une information supplémentaire dans un canal adjacent au signal audio destiné soit à l'auditeur soit à une application cible. L'existence de cette information n'est donc pas cachée. Elle n'est pas non plus nécessairement liée au contenu du signal audio. Cette subtilité conduit certains auteurs à créer une distinction entre la transmission de données et le tatouage [CMB02]. Pourtant, ces deux classes applicatives présentent de nombreux points communs, tant sur les techniques utilisées que sur les contraintes de l'application. Nous n'effectuerons donc pas de distinction de nomenclature entre tatouage et transmission de données.

Le **contexte de cette thèse** est celui du **tatouage pour la transmission de données**, certes souvent effacé devant les attentes industrielles en matière de protection des documents, mais aux enjeux non moins innovants. Insérer une information dans un signal audio trouve son intérêt dans toutes les applications [BMGL96] pour lesquelles on ne dispose pas de bande passante supplémentaire (les transmissions analogiques par exemple) et on ne peut garantir le format du signal (l'information, si elle est insérée dans l'en-tête du fichier audio par exemple, ne sera pas conservée après conversion sous-format analogique).

L'engouement pour le domaine, depuis le milieu des années 90, a suscité de nombreux débats [Bar03a, Bar03b], ayant trait à l'efficacité des techniques de tatouage [1]. Ils ont fortement incité les chercheurs à établir avec précision les applications du tatouage (qu'elles soient à but sécuritaire ou de transmission de données) et à formaliser les attentes, les contraintes et les limites de ces applications. Les sections suivantes synthétisent ces différents points dans le cas du tatouage pour la transmission de données.

1.2 Applications

Les applications attenantes aux techniques de tatouage pour la transmission de données peuvent permettre [GCG+03, Mou05, CWL01] :

- ***l'annotation de documents sonores*** pouvant servir d'aide à l'indexation dans les bases

1. Ces débats ont même conduit certains [Her02] à tenter de démontrer que le tatouage n'est pas la solution attendue par les industriels en matière de protection. L'argument avancé est le suivant : étant donné la multitude de traitements qui peuvent être effectués sur les documents numériques, aucune méthode de dissimulation d'information ne sera suffisamment robuste pour garantir une signature ineffaçable. Les chercheurs ont répondu à cette attaque en explicitant l'attente en matière de robustesse : la technique d'enfouissement de données doit être robuste aux traitements licites ou illicites avec un niveau de robustesse tel que le coût de l'effacement de la signature devient supérieur au coût de l'utilisation légale du document.

de données : l'information cachée à destination de l'auditeur ou de l'administrateur de la base de données peut indiquer le nom de l'artiste, le lieu de l'enregistrement ou toute autre donnée relative au signal. Cette annotation peut être faite en studio ou en temps-réel lors d'un concert [Tac03].

- ***l'amélioration de systèmes de transmission existants*** : l'information peut par exemple permettre l'écoute stéréo dans un système de transmission initialement dédié aux signaux mono ou diffuser le signal audio sous format numérique caché dans la bande FM analogique [CS00].

- ***l'identification du document*** : Cette application, souvent associée au tatouage de protection, peut être envisagée dans un contexte de transmission de données dès lors que l'identifiant ne sert pas de preuve de propriété mais permet d'établir des statistiques sur l'utilisation du document. L'information identifie le signal audio lors de son transfert dans un réseau de diffusion (radio, télévision ou Internet). Elle permet, par exemple, de savoir sur quelle radio a été diffusé le signal audio, combien de fois, et à quelle heure [NTK02]. Ces statistiques peuvent intéresser notamment les analyseurs de marché. Cette identification peut également remplacer les systèmes de mesure d'audience (établis pour des transferts analogiques alors que les diffusions télévisées tendent à devenir numériques). Cette application a d'ailleurs fait l'objet du projet RNRT AQUAFLUX [2] [Rés01]. Elle est également envisagée pour remplacer les identifications vocales des pilotes lors de leurs échanges avec les bases de contrôle [3] [HHKK04].

- ***le contrôle d'applications cibles*** : L'information peut être destinée à une application adjacente à l'écoute du signal audio, nécessitant la mise en œuvre d'un décodeur spécifique. Deux exemples peuvent être cités. Le premier fait référence à certaines stations de radio FM [CMB02], qui ont utilisé l'information cachée pour réduire le bruit lié à la transmission FM du signal audio. Le second est le projet RNRT ARTUS [Rés03] dont l'enjeu est d'utiliser le tatouage de la séquence audio visuelle pour transmettre des informations de mouvement à un clone virtuel. Ce clone, synthétisé par un décodeur spécifique, sera incrusté dans l'image télévisée pour reproduire l'information du télétexte en langage des signes.

- ***l'ajout d'informations publicitaires*** : le signal audio peut cacher des informations annexes à diffuser vers un auditeur. Un système de tatouage pour le télé-achat, l'IVDS [4] [TB96] a d'ailleurs été proposé : l'information cachée contient la référence du produit mis en vente par le télé-achat.

1.3 Objectifs et contraintes

Le système de tatouage, s'il est destiné à une application d'augmentation de contenu, doit satisfaire un ensemble d'objectifs et de contraintes spécifiques. On dénombre :

2. AQUA marquage de FLUX multimédia (audio et vidéo) pour leur transmission, diffusion, circulation, distribution sur les réseaux hétérogènes de télécommunications

3. En effet, le contrôle aérien régulant le traffic s'effectue par l'intermédiaire d'une fréquence radio, partagée par tous les pilotes des avions en vol : plusieurs échanges entre différents pilotes et la station de contrôle peuvent avoir lieu en même temps. Il est alors primordial que le pilote communique, à chaque échange, son identité. Plutôt que d'être effectuée vocalement, cette identification pourrait être réalisée par tatouage de la communication radio.

4. Interactive Video and Data Services

- l' ***inaudibilité*** du tatouage. L'information tatouée ne doit pas dégrader perceptuellement le signal audio dans lequel elle est insérée ; son insertion doit être transparente.

- le ***débit*** et la ***fiabilité de transmission***. La vocation du système étant de transmettre une information via un signal audio, elle doit permettre au même titre que les chaînes de communication classiques une fiabilité de transmission aussi élevée que possible pour un débit de transmission aussi grand que possible.

- la ***robustesse***. Transmettre une information cachée dans le signal audio suppose que le système mis en œuvre soit robuste aux perturbations introduites par l'ensemble des canaux de diffusion au travers desquels le signal audio peut être véhiculé : la fiabilité de transmission doit être maintenue si une dégradation affecte le signal audio tatoué. Contrairement au tatouage de protection, ces perturbations sont exemptes de la notion de pirate. L'information tatouée ayant vocation d'augmenter le contenu à destination de l'utilisateur ou d'une application cible, un pirate n'aurait rien à gagner à empêcher sa détection. Les perturbations à envisager correspondent donc à l'ensemble des dégradations licites qui peuvent être effectuées sur un signal audio lors de sa production en studio ou lors de son transfert dans un réseau de diffusion [Phi, Tel]. Ces perturbations, même exemptes des perturbations illicites du pirate, sont nombreuses : elles comprennent les modifications de format du signal (notamment la compression MPEG), de dynamique, de l'échelle des temps ou les opérations de filtrage. Le choix de ces perturbations et de leurs paramètres est limité : seul celles induisant une dégradation perceptuelle acceptable sont considérées. En effet, si la dégradation fait perdre au signal audio sa valeur (sa qualité perceptuelle), on ne peut alors exiger de l'information cachée d'être correctement détectée [5].

- la ***capacité d'insertion***, c'est à dire la quantité maximale d'information pouvant être transmise pour une probabilité d'erreur quasi nulle. Le débit d'insertion exigé par le système de tatouage est fortement dépendant de l'application envisagée (d'une centaine de bits pour l'animation du clone dans le projet ARTUS à un millier dans le cas d'applications pour la réduction de bruit). Dans les deux cas, il convient de choisir un système de tatouage permettant une capacité d'insertion la plus grande possible.

- la ***détection du tatouage aveugle***. La détection de l'information cachée doit être effectuée directement à partir du signal audio tatoué, sans avoir connaissance du signal audio original.

- le ***coût***. Certaines applications étant destinées à fonctionner en temps-réel, le coût en terme de temps de calcul peut être un critère déterminant dans le choix d'une technique de tatouage.

Satisfaire ces objectifs définit la notion de performance d'un système de tatouage. Différents critères d'évaluation doivent alors être mis en œuvre pour juger de la performance de ce système. Le chapitre qui suit en fait état.

5. Excepté dans le cas particulier d'applications qui permettraient la reconstruction du signal audio (son débruitage par exemple)

Chapitre 2

Critères et mesures de performances d'un système de tatouage

Etablir les performances d'un système de tatouage est directement lié aux quatre contraintes imposées au système, que sont l'inaudibilité de l'information cachée, la transmission la plus fiable possible pour un débit de transmission le plus élevé possible, la robustesse de la transmission à un ensemble de perturbations et la mise en œuvre peu coûteuse en temps de calcul, en vue d'applications temps-réel.

Différents critères de performance, évaluant ces quatre contraintes, doivent donc être mis en œuvre. Ce chapitre fait état des mesures proposées par la littérature et des choix que nous avons effectués pour évaluer les performances de notre système.

2.1 Mesurer la transparence du tatouage

Garantir la transparence du tatouage est sans doute la contrainte première d'un système de tatouage. Le signal audio tatoué doit être identique au signal audio original d'un point de vue perceptif. Seul un ensemble de tests d'écoute réalisés par un grand nombre d'auditeurs peut établir de manière fiable la transparence du tatouage.

Ils s'avèrent malheureusement longs et coûteux surtout en phase de conception du système. Il convient alors de disposer d'une mesure objective de la différence perceptuelle de deux signaux, qui permette d'évaluer la performance du système pendant la phase de conception. Ces différentes mesures sont présentées dans les sections qui suivent.

2.1.1 Mesures subjectives de la transparence du tatouage

La première catégorie de tests d'écoutes permet de décider si le tatouage appartient à l'une des deux catégories : celle des tatouages audibles et celle des tatouages transparents.

Une première mise en œuvre est celle d'Arnold présentée dans [Arn02] sous l'appellation *two-alternative-forced-choice* [1]. Le test est précédé d'une phase d'entraînement pendant laquelle l'auditeur compare le signal original et sa version tatouée jusqu'à penser percevoir leurs différences.

1. choix forcé à deux alternatives

Notes	SDG	Qualificatif de la dégradation
5	0	*Imperceptible*
4	-1	*Perceptible mais non gênante*
3	-2	*Légèrement gênante*
2	-3	*Gênante*
1	-4	*Très gênante*

Table 2.1 – Échelle de dégradation à cinq notes et valeurs de la SDG associée.

L'auditeur est ensuite confronté à N séries d'écoute. Chaque série comprend quatre paires de signaux, jouées de manière aléatoire : (x, x), (x, y), (y, x) et (y, y), où x désigne le signal audio original et y le signal audio tatoué. Il est alors demandé à l'auditeur si les deux signaux entendus sont identiques ou non. S'il se trompe dans 50% des cas, le tatouage est transparent.

Une seconde méthode est proposée dans la recommandation UIT [2]-R BS. 562 [56299]. Son principe est le suivant : à chaque écoute, l'auditeur se voit proposer 3 sources sonores (A, B, C). La source A est toujours le signal audio original x. Les sources B et C sont affectées aléatoirement selon l'écoute : soit B est le signal audio x et C le signal audio tatoué y, soit à l'inverse, B est y et C est x. L'auditeur doit alors décider qui de B ou de C est le signal audio tatoué y. Là encore, si l'auditeur se trompe dans 50% des cas, le tatouage est transparent.

Cette seconde méthode ne s'arrête pas à la simple question de décider si le tatouage est ou non transparent. Elle propose en plus de quantifier la différence perceptuelle entre le signal audio tatoué y et le signal original x, autrement dit la dégradation perceptuelle introduite par le tatouage. Pour juger cette dégradation, une échelle à cinq notes est proposée par la recommandation UIT-R BS. 562, chaque note qualifiant un degré de dégradation. Cette échelle est présentée table 2.1. Après avoir décidé qui de B ou de C est le signal testé, on demande à l'auditeur d'évaluer les dégradations en B par rapport à A, et en C par rapport à A sur l'échelle de dégradations à 5 notes. Des deux notes obtenues est calculé un unique attribut, traduisant la qualité perceptuelle du signal tatoué et intégrant toutes les différences détectées entre x et y. Cet attribut est appelé la ***note différentielle subjective*** et noté SDG [3]. Elle est définie par :

$$\mathrm{SDG} = \mathrm{Note}(y) - \mathrm{Note}(x). \tag{2.1}$$

Les valeurs de SDG varient sur une seconde échelle d'évaluation, également présentée table 2.1. Cette seconde échelle possède les même qualificatifs de dégradation que l'échelle de dégradation à 5 notes : 0 correspond à une dégradation imperceptible et -4 à une dégradation jugée comme très gênante.

2.1.2 Mesures objectives

Pour permettre de mesurer la différence perceptuelle entre les signaux lors de la conception d'un système de tatouage sans mettre en œuvre des tests d'écoute coûteux, différentes mesures objectives ont été proposées. La majorité de ces mesures évalue la distorsion en analysant la différence entre le signal tatoué y et le signal audio original x [KP99] : le signal de tatouage (indépendamment de la méthode d'insertion) est ici défini par $t = y - x$. Leur simplicité de mise

2. Union Internationale des Télécommunications
3. pour l'anglais Subjective Difference Grade

en œuvre les rend très populaires. Parmi elles, les deux mesures majoritairement utilisées en audio sont le rapport de puissance moyen entre le signal audio original x et du signal de tatouage t, appelé Signal to Noise Ratio (SNR) [GB00], et l'Erreur Quadratique Moyenne (EQM) [CW01]. Ces deux mesures sont définies par :

$$\mathrm{SNR} = \frac{1}{M}\sum_{m=0}^{M-1} 10\log_{10}\left[\frac{\sum_{n=0}^{N-1} x(mM+n)^2}{\sum_{n=0}^{N-1} t(mM+n)^2}\right] \tag{2.2}$$

$$\text{et } \mathrm{EQM} = \frac{1}{M}\sum_{m=0}^{M-1}\left[\frac{1}{N}\sum_{n=0}^{N-1} t(mM+n)^2\right], \tag{2.3}$$

où N est la taille des fenêtres d'analyse des signaux $x(n)$ et $t(n)$ et M le nombre de fenêtres d'analyse.

Bien que ces mesures de distorsion soient directement liées à la puissance du tatouage, elles sont souvent jugées inadéquates pour la mesure effective de la qualité auditive. En effet, elles ne prennent pas en compte les particularités du système de perception auditif humain et notamment les phénomènes de masquage. D'autres mesures, basées sur un seuil de perception modélisant le système auditif, ont donc été proposées. Ce seuil, appelé seuil de masquage et détaillé section 5.1.1.2, traduit la limite fréquentielle à satisfaire par le spectre du signal de tatouage pour être inaudible en présence du signal audio. Un rapport tatouage à masque, appelé Noise to Mask Ratio (NMR) et actualisé toutes les fenêtres d'analyse, peut alors être défini [AS02] :

$$\mathrm{NMR}(m) = 10\log_{10}\left(\frac{1}{N}\sum_{n=0}^{N-1} t(mM+n)^2\right) - 10\log_{10}\left(\sigma_M^2(m)\right), \tag{2.4}$$

où $\sigma_M^2(m)$ est la puissance du seuil de masquage sur la m-ième fenêtre d'analyse. Ce critère ne prend malheureusement en compte qu'une mesure globale du seuil et ne permet pas d'évaluer les distorsions locales introduites par le tatouage.

Il paraît donc nécessaire de disposer d'une mesure objective fiable, qui intègre l'ensemble des caractéristiques du système auditif humain. Ce problème, déjà rencontré lors de l'évaluation des algorithmes de compression audio, a permis la conception et la normalisation d'un algorithme de mesure, l'algorithme PEAQ [Uni01], qui peut être utilisé dans un contexte de tatouage. Cet algorithme ne prétend pas remplacer les tests d'écoutes mais permet d'obtenir facilement une mesure relativement fiable de la transparence du tatouage, ce qui nous a naturellement conduit à le choisir comme critère d'évaluation des performances.

2.1.3 L'algorithme PEAQ

L'algorithme PEAQ est la synthèse de 6 méthodes[4] développées dans les années 90 avec l'objectif de fournir une mesure objective de la qualité sonore d'un signal audio comprimé. Sa mise en œuvre est le résultat d'une normalisation par l'UIT-R [Uni01].

4. Ces 6 méthodes étant :
- l'indice de perturbation DIX (Distorsion IndeX) [TK96],
- le rapport bruit à masque (NMR) [Bra87],
- le système de mesure OASE (Objective Audio Signal Evaluation) [Spo97],
- la mesure perceptuelle de la qualité du son PAQM (Perceptual Audio Quality Measure) [BS94],
- le système PERCEVAL (PERCeptual EVALuation of the quality of audio signals) [PMMeS92],
- la mesure perceptuelle objective POM (Perceptual Objective Measurement) [Col94] et
- la Toolbox Approach.

Les tests d'écoute proposés par la recommandation UIT-R BS. 562 présentés précédemment quantifient la distorsion introduite par le tatouage sous la forme d'une note, la SDG, associée à un attribut de dégradation. L'objectif de PEAQ est d'estimer cette note à partir de mesures objectives : il fournit donc une seconde note, l'***ODG***[5], révélant la qualité du signal tatoué de façon suffisamment fiable pour s'affranchir de tests d'écoute subjectifs pendant la conception de l'algorithme de tatouage.

Pour quantifier la qualité du signal tatoué par rapport à celle du signal audio, l'algorithme PEAQ intègre un ensemble de mesures qui établissent les différences d'excitations provoquées par les deux signaux sur le système auditif. Ces mesures relèvent aussi bien de caractéristiques physiques[6] que de considérations psychoacoustiques. Ces dernières utilisent un modèle auditif (modélisant le comportement physiologique de l'oreille) : ce modèle transpose les caractéristiques spectrales des signaux (en Hertz) en caractéristiques d'excitation sur la membrane basilaire de l'oreille interne (en Bark), siège de la détection et de l'analyse d'un son. Les caractéristiques d'excitation sont ensuite traitées pour fournir différents critères de comparaison : les caractéristiques de masquage, aussi bien temporelles que fréquentielles (via un seuil de masquage), l'intensité de perception et les caractéristiques de modulation de l'enveloppe des signaux. Différentes mesures psychoacoustiques en sont déduites[7]. Les mesures physiques et psychoacoustiques sont finalement combinées dans un réseau de neurones artificiel pour obtenir l'ODG. Ce réseau de neurones est spécifiquement établi de sorte que la valeur de l'ODG qui en résulte coïncide avec la valeur de la SDG (obtenue à la suite de tests d'écoute).

Il faudra être vigilant concernant l'utilisation de l'algorithme PEAQ pour les systèmes de tatouage. En effet, cet algorithme a été établi spécifiquement dans le cadre des codeurs audio. La méthode de mesure des caractéristiques d'excitation étant indépendante de l'application, son utilisation pour le tatouage n'est pas remise en cause. Par contre, le réseau de neurones est lui spécifiquement établi pour les distorsions introduites par la compression audio ; il risque d'être moins adapté aux distorsions introduites par le tatouage qu'il ne pourrait l'être. Une étude devrait donc être effectuée pour vérifier la concordance des ODGs et des SDGs obtenus pour les signaux de tatouage et pour modifier au besoin le réseau de neurones ; mais elle s'inscrit en dehors du contexte de cette thèse.

2.2 Évaluer la fiabilité et la robustesse de la transmission

2.2.1 Taux d'erreur binaire

L'information à tatouer est supposée codée sous la forme d'une séquence de bit. La fiabilité de transmission du tatouage est donc directement liée à la probabilité d'erreur de transmission P_e. Cette probabilité d'erreur peut être estimée par le calcul du Taux d'Erreur Binaire (TEB). Ce TEB est le ratio entre le nombre de bits erronés et le nombre de bits total B émis. Cet estimateur de P_e est efficace [Cha96] et offre une précision ΔP_e pour un taux de confiance à 70% de :

$$\Delta P_e = \sqrt{\frac{\mathrm{TEB}(1-\mathrm{TEB})}{B}}. \tag{2.5}$$

5. Objective Difference Grade
6. mesure de la largeur de bande des signaux, structure harmonique du signal de tatouage notamment.
7. moyenne de l'intensité acoustique, moyenne des différences de modulation, rapport tatouage à masque (NMR) et probabilité de détection des composantes fréquentielles étant donné le seuil de masquage.

Notre système étant dédié à la transmission d'une information via le canal audio, la fiabilité de transmission du système devra donc être évaluée en fonction du débit de transmission R.

2.2.2 Perturbations

Les perturbations, auxquelles un système de tatouage dédié à l'augmentation de contenu se doit d'être robuste, sont l'ensemble des opérations licites qui peuvent être appliquées au signal audio tatoué ; comme nous l'avons déjà précisé, dans notre contexte, les perturbations volontaires qu'un pirate mettrait en œuvre pour supprimer le tatouage ne sont pas à prendre en compte.

Les outils de traitement des signaux audio (CoolEdit, Adobe Audition, Soundforge ou les logiciels professionnels pour les traitements en studio) offrent de nombreuses possibilités, tant sur les post-traitements des signaux audio que sur les paramètres de ces traitements. Différentes études ont été menées pour répertorier l'ensemble des perturbations à considérer. Grâce au challenge proposé par le SDMI [8] [SDM, CWLF01] ou à la mise en place de l'outil Stirmark [Lan] [9], on dispose actuellement d'une liste relativement exhaustive [10]. De cette liste, nous ne conserverons que les perturbations induisant une dégradation limitée de la qualité auditive du signal audio : en effet, la dégradation doit conserver en grande partie la valeur (en terme de qualité sonore) du signal audio. Plusieurs catégories peuvent être répertoriées [SPR+01] :

- les ***attaques de la dynamique***. Ces attaques consistent à modifier la sonie du signal audio : il s'agit de l'augmenter et de la diminuer dans la plupart des cas ou de la soumettre à des transformations non linéaires (expanseur ou compresseur de dynamique, amplification des basses, lissage du signal, etc...).

- les ***attaques par filtrage***, utilisées pour augmenter ou supprimer certaines parties du spectre du signal audio : filtrage passe-bas, passe-haut, tout-pôle mais également égaliseur de spectre, modification de la répartition du signal entre la voie de gauche et la voie de droite dans le cas de signaux stéréo sont autant de configurations possibles.

- les ***modifications d'ambiance***, qui permettent de simuler les effets d'une pièce lors de l'écoute, notamment les ajouts de réverbération (ou écho). Ces perturbations modifiant la perception du son, elles ne seront pas prises en compte dans cette étude.

- les ***changements de format*** : cette catégorie regroupe le rééchantillonnage ou l'interpolation des signaux à une fréquence d'échantillonnage différente du signal original, la modification du nombre de bits utilisés pour la quantification des échantillons, les compressions du signal (compression MPEG par exemple), les conversions sous format analogique (CNA-CAN [11]) mais également des problèmes dits de changement de la ***polarité*** du signal audio. Changer la polarité consiste à inverser le signe de tous les échantillons du signal audio. Ce problème se produit fréquemment lors de l'acquisition d'un signal audio analogique.

- l'***ajout de bruit***, lors d'un enregistrement d'un signal audio tatoué analogique dans un environnement bruité.

8. Secure Digital Music Initiative
9. Cet outil est devenu un standard dans le domaine du tatouage des images, il tend à le devenir dans le domaine de l'audio.
10. Cette liste prend également en compte les perturbations introduites par l'action d'un pirate, que nous ne présenterons pas ici.
11. Conversion Numérique Analogique, Conversion Analogique Numérique

- les ***effets de modulation***, parmi lesquels on distingue les vibratos, les chorus ou les modulations d'amplitude. Ces perturbations sont spécifiquement utilisées en studio pour modifier les perceptions sonores du signal audio. Elles ne seront donc pas considérées ici, puisque la qualité du signal est modifiée.

- les ***modifications de l'échelle temporelle et du pitch***. Les modifications de l'échelle des temps permettent de modifier la durée d'un signal en la dilatant ou la contractant. Elles peuvent être involontaires dans le cas d'une conversion analogique-numérique du signal audio tatoué ou volontaires pour des diffusions radios où l'extrait sonore, préalablement tatoué, doit être diffusé dans un temps imparti. Dans le cas des conversions analogiques-numériques, le signal audio tatoué peut être rééchantillonné à une fréquence d'échantillonnage légèrement différente de l'émission (inférieure à une dizaine d'échantillons) ; cette différence d'échantillonnage est souvent liée aux imperfections des quartz des convertisseurs. Dans le cas des diffusions radio, le "*time stretching*" peut être utilisé pour modifier à l'occasion la durée du signal sans modifier le pitch de sorte à satisfaire la contrainte de temps de diffusion sans altérer la perception du signal. A ces modifications de l'échelle des temps, s'ajoutent les modifications de pitch, qui permettent par exemple de modifier la hauteur de la voix ; puisqu'elles dénaturent la perception du son, ces modifications de pitch ne seront donc pas considérées dans cette étude.

2.2.3 Outils d'évaluation

Face à la multitude de perturbations et de paramétrages possibles, il devient nécessaire de disposer d'un outil d'évaluation "normalisé", afin de comparer les performances des systèmes de la littérature. Lang propose un outil exécutable [Lan], le Stirmark Audio, qui tend à devenir une référence en matière de procédure d'évaluation de la robustesse des systèmes de tatouage audio. Cet outil génère l'ensemble des versions dégradées d'un signal audio par différentes perturbations, aussi bien licites qu'illicites. Ces perturbations offrent un premier éventail de la liste des perturbations envisagées. Notre système étant exempté de la notion de pirate, nous n'extrairons de cet outil que les perturbations licites.

Certaines catégories de perturbations ne sont pas encore proposées par l'implémentation du Stirmark, notamment les compressions MPEG ou les opérations de modification d'échelle. Elles devront donc être ajoutées au protocole expérimental pour fournir une mesure significative de la robustesse des systèmes.

2.3 Mesurer le coût en temps de calcul

Le coût en temps de calcul peut intervenir dans la construction d'un système de tatouage. En effet, si le système est destiné à être temps-réel, il pourra orienter les choix vers des techniques de tatouage peu coûteuses en calcul. Le cas échéant, il peut également départager deux stratégies de tatouage dont les performances en terme de fiabilité de transmission et de robustesse sont identiques. Deux mesures peuvent permettre d'évaluer le coût :

- le ratio entre le temps nécessaire à la simulation et la durée du signal audio traité. Cette mesure est dépendante du processeur utilisé.

Signaux	Nom	Descriptions
1	*"Jonasz"*	Voix chantée masculine
2	*"Beethoven"*	Orchestre symphonique
3	*"Partita_no3"*	Violon
4	*"Gainsbourg"*	Voix chantée masculine
5	*"Mahler_symph5"*	Orchestre symphonique
6	*"Vivaldi"*	Orchestre symphonique
7	*"TuckPatti_AsTimeGoesBy"*	Voix chantées féminines
8	*"Artus_2_44"*	Voix parlée masculine avec bruits de fond
9	*"Angelina"*	Voix chantée masculine
10	*"Camille"*	Orchestre symphonique
11	*"Dinah"*	Voix chantée masculine
12	*"Renaud"*	Voix chantée masculine
13	*"Preacher"*	Voix chantée féminine
14	*"Prokofiev"*	Orchestre symphonique
15	*"Khachaturian"*	Orchestre symphonique
16	*"Fetes"*	Orchestre symphonique
17	*"Cranberries"*	Voix chantée féminine
18	*"DePalmas"*	Voix chantée masculine
19	*"Rembrandts"*	Voix chantées masculines
20	*"Goldeneye"*	Voix chantée féminine

Table 2.2 – Signaux audio du corpus.

- le nombre d'instruction par seconde (MIPS [12]) ou le nombre de multiplications et d'accumulations par seconde (puisqu'en traitement du signal, les instructions correspondent essentiellement à des produits scalaires).

2.4 Protocole expérimental

Les performances des systèmes de tatouage, présentés dans ce manuscrit, sont évaluées, sauf mention contraire, par le protocole de mesures suivant. Un corpus de 20 signaux audio mono, décrit table 2.2, est considéré : ces signaux sont échantillonnés en qualité CD (à une fréquence d'échantillonnage $F_e = 44,1$ kHz) et sont de style divers allant d'extraits d'orchestres symphoniques à des voix parlées en passant par des extraits de solo et des voix chantées.

Les critères de performances considérés sont les suivants :

- ***l'inaudibilité de l'insertion du tatouage***. Cette inaudibilité est évaluée grâce à l'outil d'évaluation PEAQ décrit précédemment et par des tests d'écoute subjectifs réalisés de manière informelle par trois auditeurs. La mesure de l'ODG moyenne obtenue avec tatouage des 20 signaux de la base test par une séquence de 5000 bits chacun à un débit de 206 bit/s [13] sera proposée.

12. Million Instructions Per Second
13. Ce débit est choisi de manière arbitraire dans la mesure où l'ODG est quasiment indépendante du débit de transmission pour le système de tatouage qui sera établi par la suite.

Catégories	Perturbations	Paramètres
Attaques de la dynamique	Modification de la sonie*	Multiplication de l'amplitude par $0,5$
	Compresseur de dynamique	$2:1$
Attaques par filtrage	Filtrage passe-haut*	Coupure à 200 Hz
	Filtrage passe-bas*	Coupure à 9 kHz
Modifications d'ambiance	Echo*	Ajout d'une copie atténuée du signal (10%) avec 400 ms de retard
Ajout de bruit	Ajout de bruit blanc	Rapport Signal à Bruit (RSB) 30 dB
Changements de format	Compression MPEG	Débit de 96 kbit/s et 64 kbit/s
	Rééchantillonage*	De $44,1$ kHz à $29,4$ kHz
	Quantification	Échantillons quantifiés sur 12 bits
Modifications de l'échelle temporelle	*Time stretching*	Simulé par un rééchantillonnage à une fréquence variable

Table 2.3 – Perturbations prises en compte pour évaluer la robustesse du système de tatouage. Celles suivies par une * sont réalisées à l'aide de l'outil Stirmark audio.

- ***le TEB pour un canal sans perturbation***. Chaque signal du corpus est tatoué par une séquence de 5000 bits choisis aléatoirement pour différents débits de transmission R. Ce choix permet d'obtenir une mesure du TEB pour une séquence globale de $B = 100\,000$ bits avec une précision $\sqrt{\text{TEB}(1-\text{TEB})/B}$ (conformément à l'équation 2.5) pour un taux de confiance de 70%. Au regard des TEBs obtenus par la suite par le système de tatouage, cette précision est de l'ordre de 10^{-3}. Les valeurs du TEB inférieures à 10^{-3} ne sont donc pas significatives. Les débits de transmission considérés pour simuler la chaîne de tatouage sont donc choisis de sorte à ce que les TEBs soient de l'ordre de 10^{-3} ou supérieur à 10^{-3}. Il pourra être reproché la faible précision des mesures : chercher à atteindre un TEB de 10^{-3} quand les chaînes de communication actuelles visent des TEB de l'ordre de 10^{-6} à 10^{-9} prête à commentaires. Ce TEB constitue un compromis entre la précision des mesures et le temps de calcul pour la simulation. En effet, pour obtenir des TEBs de l'ordre de 10^{-6} (étant donné les performances de la chaîne), les débits de transmission sont de l'ordre de 20 bit/s ; il faudrait alors traiter plus de 10 heures de signal audio pour obtenir une précision de 10^{-6}, ce qui n'est pour l'instant pas réalisable en pratique. Un TEB de 10^{-3} pour une précision de 10^{-3} peut être obtenu à 50 bit/s en traitant 16 mn, ce qui est nettement plus raisonnable à mettre en pratique.
  Pour finir les valeurs de débits présentés R dans tout le manuscrit prennent en compte les patterns de synchronisation introduites pour les besoins du mécanisme de synchronisation, même lorsqu'il n'en est pas fait mention.

- ***la robustesse aux perturbations***. Différentes perturbations sont considérées offrant un échantillon relativement exhaustif des perturbations licites auxquelles le système de tatouage se doit d'être robuste. Ces perturbations sont énumérées table 2.3 avec leurs paramètres. Nous considérerons : une opération de compression MPEG réalisée par le codeur LAME (codeur MPEG 1 Layer 3) à différents débits de compression, des opérations de filtrage (passe-haut et passe-bas), l'ajout d'écho, le sous-échantillonnage du signal audio tatoué et la modification du nombre de bits utilisés pour la quantification des échantillons,

l'ajout de bruit blanc, la modification de l'amplitude du signal, une compression de la dynamique et une opération de désynchronisation simulant une opération de *time stretching* (en appliquant la formule d'interpolation établie section 11.1). Les opérations de filtrage et le compresseur de dynamique sont réalisés par l'outil Stirmark avec leurs paramètres par défaut, tels qu'ils sont proposés au moment du téléchargement de l'exécutable. Les autres perturbations sont codées en Matlab©.

- ***le temps de calcul***. Le temps de calcul est mesuré par le ratio entre le temps nécessaire à la simulation et la durée du signal audio traité. Cette mesure est très dépendante du logiciel de programmation (ici Matlab©et le langage C) et du processeur utilisé (un Pentium 4, $1,8$ GHz avec 512 Ko de RAM). Elle sera donnée principalement à titre indicatif ou à titre comparatif, puisque l'implémentation des systèmes n'a pas été optimisée.

Chapitre 3

Etat de l'Art

Le tatouage est l'art d'exploiter les imperfections du système auditif humain pour transmettre une information inaudible dans un signal audio. Notre contexte applicatif s'attachant à maximiser la quantité d'information transmise (le débit de transmission) tout en maximisant la fiabilité de transmission (mesurée par le TEB), l'analogie entre les systèmes de tatouage et les chaînes de communication est immédiate : le signal de tatouage, porteur de l'information à émettre, est transmis dans le canal que constitue le signal audio. Les systèmes de tatouage sont donc conçus sous la forme d'une chaîne de communication, dont le schéma de principe est donné figure 3.1. Ce schéma reprend les trois étapes caractéristiques d'une chaîne de communication : l'émetteur, organe de construction du signal de tatouage, le canal, siège des perturbations classiquement appliquées au signal audio et le récepteur qui extrait du signal reçu l'information cachée.

Les conditions de transmission du tatouage sont nettement plus sévères que celles des canaux Bruit Blanc Additif Gaussien (BBAG) étudiés par la théorie des communications numériques. Ces conditions sont liées aux particularités du canal de transmission audio. Ce canal est :

- très corrélé, nécessitant la mise en œuvre d'outils de décorrélation,
- non-stationnaire, empêchant l'utilisation de procédures d'égalisation coûteuses en observations,
- de rapport signal à bruit très défavorable. En effet, pour être inaudible, le tatouage doit être choisi de faible puissance devant le signal audio.

De fait, le tatouage doit faire appel à des techniques de communications spécifiques (modulation par étalement de spectre, codes correcteurs d'erreur, procédures d'égalisation adaptées) pour permettre la détection correcte de l'information cachée. Les recherches se sont donc initialement concentrées sur la mise en œuvre de ces techniques pour proposer des systèmes de tatouage efficaces. Malheureusement leurs performances, notamment en terme de capacité d'insertion, sont inévitablement limitées par le bruit audio.

En 1999, Cox *et al.* [CMM99] apportent un souffle nouveau au domaine en établissant l'analogie entre le tatouage et les chaînes de communication avec information adjacente : les systèmes de tatouage présentent en effet la particularité de connaître à l'émetteur le bruit de canal audio. Les résultats théoriques de Costa [Cos83], établis pour une configuration particulière de chaîne de communication avec information adjacente, ont montré que les performances du système pouvaient être rendues indépendantes du signal audio : plutôt que de lutter contre le bruit audio, l'émetteur doit utiliser ce signal pour construire un tatouage adapté garantissant la détection robuste de l'information tout en conservant un caractère inaudible. Une seconde génération de

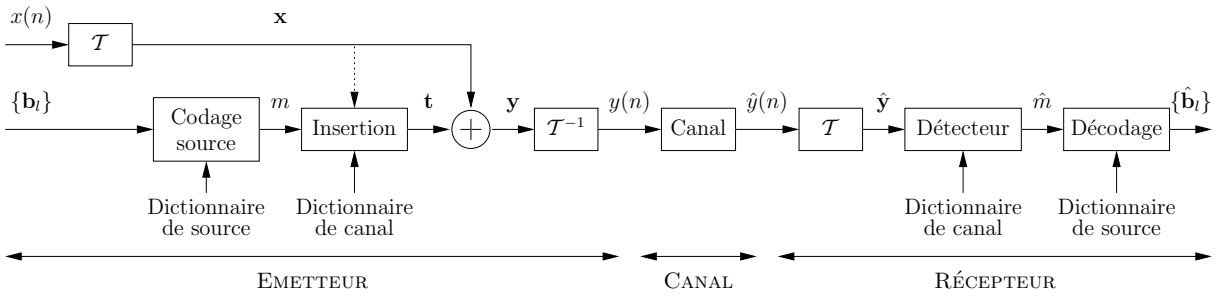


Figure 3.1 – Schéma de principe d'un système de tatouage.

systèmes s'est donc construite autour de ce modèle. Ces systèmes sont alors qualifiés d'informés parce qu'ils exploitent la connaissance *a priori* du signal audio à l'émetteur. La technique d'enfouissement de données relève alors d'une stratégie : celle de choisir un tatouage optimal garantissant l'inaudibilité de la dégradation introduite par son insertion et la maximisation de la fiabilité de transmission. De nombreuses études, s'attachant notamment aux performances théoriques de ces systèmes, ont permis d'établir l'efficacité des stratégies informées sur les stratégies non informées. Pourtant, la mise en œuvre pratique de ces stratégies est encore d'actualité, puisque l'on ne dispose toujours pas d'un système de tatouage atteignant les performances théoriques.

Les contributions dans le domaine du tatouage que ce soit pour des images, des vidéo ou des signaux audio sont de plus en plus nombreuses ; les techniques de tatouage étant souvent adaptées au support de transmission (les contraintes imposées au tatouage n'étant pas les mêmes lorsqu'il s'agit de cacher visuellement un message et de le cacher auditivement), nous nous sommes principalement intéressés aux travaux spécifiquement dédiés à l'audio. Les contributions génériques (c'est à dire ne dépendant pas du signal support) seront également présentées. La première section 3.1 propose une synthèse des travaux dans le domaine du tatouage audio et permet d'établir un modèle général des systèmes de tatouage. La seconde section 3.2 met en évidence les axes de recherche les plus récents pour les systèmes non informés et en donne un ordre de grandeur des performances de ces systèmes. La troisième section 3.3 est dédiée aux systèmes informés, cœur de ce travail de thèse. Elle établit un état des recherches en matière de stratégie d'insertion informée en mettant en évidence le gain offert par ces techniques, leurs limites et les enjeux restant à atteindre. La dernière section 3.4 s'intéresse au problème de la désynchronisation dans les systèmes de tatouage, qui compte parmi les perturbations les plus dégradantes pour un système de tatouage. Ce problème nécessite de recourir à des mécanismes spécifiques dont la mise en œuvre peut être dissociée de la stratégie d'insertion.

3.1 Modèle général d'un système de tatouage

Dans un contexte applicatif de transmission d'une information cachée dans un signal audio, l'analogie entre système de tatouage et chaîne de communication, avec ou sans information adjacente, est immédiate. Le schéma de principe, présenté figure 3.1, reprend les trois étapes classiques d'une chaîne de communication : l'émetteur, où le signal de tatouage est construit, le canal siège de perturbations classiquement appliquées au signal audio et le récepteur, qui extrait l'information cachée du signal audio reçu.

3.1.1 Schéma de principe

L'émetteur vise à insérer de manière imperceptible une information binaire, séquence de L bits b_l, au signal audio $x(n)$. Cette insertion est effectuée dans un ***domaine d'insertion*** approprié : une transformation $\mathcal{T}$ extrait du signal audio $x(n)$ un vecteur de N_s caractéristiques, noté $\mathbf{x}$ et également appelé ***signal hôte***.

L'information binaire est interprétée par une étape de *codage de source* qui associe à chaque K-uplet binaire un représentant m parmi M. Ce représentant est associé de manière unique aux bits émis et est choisi dans un ***dictionnaire de source***. Il est ensuite inséré dans le signal audio à l'aide d'une ***fonction d'insertion*** f. Cette fonction d'insertion construit dans un premier temps le vecteur de tatouage $\mathbf{t}$, homogène aux caractéristiques du signal audio $\mathbf{x}$, à l'aide d'un ***dictionnaire de canal*** qui réalise l'association bijective entre le vecteur de tatouage $\mathbf{t}$ et le représentant m. Elle insère dans un deuxième temps le vecteur de tatouage $\mathbf{t}$ (souvent en l'ajoutant) au vecteur audio $\mathbf{x}$. Le débit de transmission est alors de $R = K$ bits par vecteur.

L'ensemble de ces deux étapes - codage de source et insertion - définit la stratégie d'insertion. Lorsqu'elle prend en compte la connaissance *a priori* du signal audio, elle devient informée.

La stratégie d'insertion conduit finalement au signal audio tatoué $y(n)$. Ce signal est obtenu par transformation inverse du vecteur de caractéristiques $\mathbf{y}$ défini dans le domaine d'insertion vers le domaine temporel. Le signal de tatouage $t(n)$ est alors défini comme la différence entre le signal audio tatoué $y(n)$ et le signal audio original $x(n)$:

$$y(n) = x(n) + t(n). \tag{3.1}$$

Le récepteur traite le signal audio reçu $\hat{y}(n)$, *i.e.* $y(n)$ éventuellement détérioré par le canal de transmission, en le projetant dans l'espace d'insertion. Une étape de détection, dépendante de la fonction d'insertion choisie mais indépendante du signal audio original, est mise en œuvre pour décider du représentant $\hat{m}$. Cette étape peut faire intervenir des pré-traitements qui projettent le signal reçu dans un espace de détection, différent de l'espace d'insertion. Une fois le représentant $\hat{m}$ décidé, le codage de canal est ensuite inversé, donnant lieu à l'information binaire reçue $\hat{b}$.

3.1.1.1 Domaines d'insertion et de détection

Le choix des domaines d'insertion et de détection est une étape délicate dans la conception du système de tatouage. Les performances du système, notamment en terme de fiabilité de transmission, en sont fortement dépendantes : en effet, le TEB dépend des caractéristiques du bruit audio $\mathbf{x}$ projeté dans ce domaine. Différents critères régissent le choix d'un domaine adapté [AWS03, CMB02]. Ce domaine peut :

- permettre de blanchir le signal audio, voire de décorréler le signal audio du signal de tatouage. Cette décorrélation tente de ramener le système de tatouage à une chaîne de tatouage BBAG idéale, dont les performances sont optimales.

- faciliter l'insertion inaudible du tatouage. En effet, permettre une altération importante de certaines composantes du signal audio sans modifier la perception de celui-ci peut faciliter la détection du tatouage.

- être invariant à certaines perturbations subies par le signal audio ; ce critère est fréquemment utilisé lorsque les perturbations désynchronisantes sont considérées. Cette

invariance facilite la conception de systèmes robustes aux perturbations.

La littérature propose différents domaines respectant au moins l'un de ces critères (mais jamais les trois). Nous avons notamment dénombré :

- le ***domaine temporel***. Ce domaine permet d'utiliser des modèles psychoacoustiques qui établissent les conditions d'inaudibilité du signal de tatouage en présence du signal audio. Dans ce domaine, l'insertion s'attache à modifier l'amplitude du signal audio [BTH96, BPN01].

- le ***domaine fréquentiel***. Ce domaine est souvent choisi pour son invariance à certaines attaques désynchronisantes : en effet, la représentation fréquentielle est robuste aux modifications de l'échelle des temps (opérations de *time stretching* à *pitch* constant) couramment appliquées aux signaux audio. Différentes transformations temps-fréquence sont proposées : la TFD [1], permettant de modifier aussi bien les coefficients du spectre d'amplitude [AS02] que la phase [BMGL96], la MCLT [2] de Malvar [KM03], la DHT [3] [SM04], l'utilisation de bancs de filtres, QMF [4] par exemple, permettant de ne sélectionner que certaines sous-bandes du spectre du signal audio [SFK02] ou la transformée en ondelettes [MT01].

- le ***domaine cepstral***. Ce domaine, très souvent considéré pour le tatouage des signaux de parole [LH00, HS02], contient l'information la plus significative de la parole : il est de fait robuste à un ensemble de traitements sur les signaux dont les opérations désynchronisantes.

- les ***représentations paramétriques*** du signal audio. L'insertion peut en effet être effectuée directement dans le bit stream MPEG du signal audio comprimé [SNHK01] ou être combinée avec l'encodeur CELP [5] [LYS05]. Elle peut également viser à modifier les coefficients de l'analyse LPC [6] [CS01] ou les paramètres de la représentation sinusoïdale du signal audio [GM04].

- les domaines résultant de la ***projection*** du signal audio dans un espace de dimensions réduites choisi pour décorréler le signal audio ou en diminuer la puissance. Ce choix est effectué soit de manière aléatoire [Che00] soit après analyse en composantes principales du signal audio [GPM04].

Aucun consensus n'a encore été établi sur le domaine d'insertion le plus approprié.

3.1.1.2 Codage de source

Différents codages de source peuvent être envisagés. Dans les cas les plus simples, ce codage consiste à associer à chaque K-uplet binaire un symbole de l'alphabet $\{0, \ldots, M-1\}$ avec $M = 2^K$. Le représentant est alors appelé ***symbole*** et le dictionnaire de source ***alphabet***.

L'encodage source peut également permettre l'introduction de codes correcteurs d'erreur : une redondance est introduite dans la séquence binaire afin d'améliorer le TEB du système. Le

1. Transformée de Fourier Discrète
2. Modulated Complex Lapped Transform
3. Discrete Hartley Transform
4. Quadrature Mirror Filter
5. Code Excited Linear Predictio
6. Linear Prediction Coding

représentant m fait alors référence à un ***mot de code*** : d'un K-uplet de la séquence binaire à tatouer est déduit un K'-uplet binaire de code. Il convient alors de distinguer ***débit de transmission utile*** et ***débit de fonctionnement***. Le débit de transmission utile est le nombre de bits d'information effectivement transmis par seconde tandis que le débit de fonctionnement est le nombre de bits (du mot de code) transmis par seconde : étant donné l'ajout de redondance, le débit utile est toujours inférieur au débit de fonctionnement. Différents codes correcteurs d'erreur sont proposés dans la littérature : les turbo codes [CTS03], les codes de Reed-Solomon [LFT02], les codes BCH [7] [SM04] pouvant exploiter un algorithme de décodage souple [Bau02], pour ne citer que quelques exemples. Ces codes correcteurs d'erreur ne sont efficaces qu'à la condition que les TEBs du système sans correction soient déjà relativement faibles. En effet, dans [CTS03], l'auteur établit explicitement le gain obtenu grâce aux turbo codes sur les TEBs d'un système de tatouage audio à débit de transmission utile. Pour de faibles débits (inférieurs à une centaine de bit/s), les TEBs obtenus par les turbo codes sont nettement inférieurs à ceux d'un système sans encodage de source ; à débits élevés, la tendance s'inverse : les turbo codes ne sont plus assez efficaces pour compenser les erreurs introduites par le débit de fonctionnement plus élevé. L'introduction de codes correcteurs d'erreur doit donc être précédée d'améliorations significatives des performances de la stratégie d'insertion au regard du débit de fonctionnement.

3.1.1.3 Insertion

La fonction d'insertion vise à construire un tatouage $\mathbf{t}$, porteur du représentant m puis à l'insérer de manière inaudible dans le signal audio $\mathbf{x}$.

Elle nécessite dans un premier temps la mise en œuvre d'une étape de ***modulation***, qui convertit le représentant m en un signal physique homogène au vecteur de caractéristiques $\mathbf{x}$. Cette étape peut être schématisée à l'aide d'un dictionnaire de canal, contenant l'ensemble des formes d'onde utilisées pour la modulation. Ces formes d'onde sont souvent des séquences pseudo-aléatoires dans le cas des systèmes additifs temporels [BTH96] ou des répliques du signal audio décalées dans le temps dans le cas des systèmes par ajout d'écho [Pet01]. La modulation associe de manière unique un vecteur $\mathbf{v}$ du dictionnaire au représentant m. Dans le cas d'un système non informé, cette association est bijective : il existe autant de formes d'onde que de représentants possibles. Dans le cas des systèmes informés, le dictionnaire peut disposer de plusieurs formes d'onde pour un même représentant m. La connaissance du signal audio est alors exploitée afin de choisir parmi ces formes d'onde, celle la plus adaptée à la transmission de m.

Garantir l'inaudibilité du tatouage fait inévitablement appel à la psychoacoustique. Cette dernière est mise à profit pour établir les conditions à satisfaire par le tatouage $\mathbf{t}$ pour garantir son inaudibilité, qui seront plus amplement détaillées en section 5.1.1. Ces conditions dépendent des caractéristiques propres du signal audio [8]. Elles s'expriment principalement sous trois formes :

- un ***facteur d'amplitude*** α, également appelé force d'incrustation. Ce facteur pondère le vecteur modulé $\mathbf{v}$ de sorte à garantir un rapport de puissance entre le tatouage résultant $\mathbf{t} = \alpha\mathbf{v}$ et le signal audio relativement faible.

- un ***seuil de masquage*** $S_M(f)$ homogène à une densité spectrale de puissance qui établit la limite fréquentielle maximale admissible à satisfaire par le tatouage. Ce seuil de masquage,

7. Bose-Chaudhuri-Hochquenghem

8. Ainsi tous les systèmes qui exploitent une analyse psychoacoustique du signal audio pourraient être qualifiés d'informés. Pourtant, ils sont plus couramment dénommés semi-informés [CMM99] car ils n'utilisent qu'une partie de la connaissance du signal audio à l'émetteur.

établi d'après un modèle psychoacoustique, régit la mise en forme spectrale du vecteur $\mathbf{v}$ dont est issu le tatouage $\mathbf{t}$. Cette mise en forme spectrale est réalisée soit dans le domaine temporel sous la forme d'une opération de filtrage [Gom02], soit dans le domaine fréquentiel sous la forme d'une pondération fréquence par fréquence du tatouage [LGP04].

- une "***modulation***" par le signal audio [Pet01, CTS03]. Le vecteur de tatouage $\mathbf{t}$ est obtenu en pondérant, caractéristique par caractéristique, le vecteur modulé $\mathbf{v}$ par le signal audio $\mathbf{x}$ atténué : $\mathbf{t} = \alpha\mathbf{v} \otimes \mathbf{x}$, où $\otimes$ désigne le produit terme à terme. Cette mise en forme du vecteur modulé permet de contrôler l'énergie du tatouage et d'éviter les phénomènes de pré-écho dans les régions du signal audio à fortes variations d'énergie [Pet01].

Par souci de simplicité, nous ne conserverons dans ce chapitre que le formalisme du facteur d'amplitude. Le tatouage obtenu s'écrit de fait : $\mathbf{t} = \alpha\mathbf{v}$.

La fonction d'insertion réalise finalement le mélange entre le vecteur audio et le tatouage $\mathbf{t}$. Elle conduit au vecteur audio tatoué $\mathbf{y}$ exprimé sous la forme :

$$\mathbf{y} = (1-\beta)\mathbf{x} + \mathbf{t} = (1-\beta)\mathbf{x} + \alpha\mathbf{v}. \tag{3.2}$$

Cette expression synthétise l'ensemble des fonctions d'insertions proposées par la littérature :

- lorsque $\beta = 0$, la fonction d'insertion, couramment utilisée dans les systèmes dit à étalement de spectre, devient ***additive*** : $\mathbf{y} = \mathbf{x} + \mathbf{t}$.

- lorsque $\beta = 1$, elle traduit une insertion ***substitutive*** : les caractéristiques du signal audio sont remplacées par celles du tatouage. Cette insertion est exploitée par de nombreuses stratégies d'insertion par quantification [Che00], mais également par des techniques de substitutions de sous-bandes fréquentielles du signal audio. Dans [BMJ95], ces sous-bandes sont prédéfinies à l'avance tandis que dans [Gar99] et [MGN02], elles sont sélectionnées en fonction du signal audio : un modèle psychoacoustique est mis à profit pour choisir les sous-bandes jugées non perceptibles. L'information du signal audio dans les sous-bandes fréquentielles sélectionnées est remplacée par une modulation BPSK [9] de l'information à émettre.

- dans les autres cas, la fonction d'insertion est qualifiée d'insertion par ***réjection*** (ici partielle) du signal audio, telle que l'a définie Chen [Che00].

3.1.1.4 Détecteur

Le module de détection est conditionné par la fonction d'insertion et le domaine d'insertion choisis. Il vise à extraire du vecteur de caractéristiques $\hat{\mathbf{y}}$, obtenu par projection du signal audio tatoué détérioré $\hat{y}(n)$ dans le domaine d'insertion, le représentant $\hat{m}$ le plus probablement émis.

Dans le cas des fonctions d'insertion additives ou des fonctions avec réjection partielle de l'hôte, le signal audio est une source d'interférences qui perturbe (parfois très fortement) la détection du tatouage. Le module de détection débute donc souvent par un pré-traitement, dont le but est de décroître l'influence du signal hôte sur le procédé d'extraction du tatouage [CTS03]. Cette étape peut prendre la forme d'une opération de blanchiment [DKL98] ou d'une procédure d'égalisation du canal [Gom02, Lar05]. Elle conduit au vecteur reçu $\hat{\mathbf{v}}$. Ce pré-traitement entraîne le module d'extraction vers un espace de détection éventuellement différent du domaine

9. Binary Phase Shift Keying

d'insertion - celui dans lequel se trouve $\hat{\mathbf{v}}$. Il nécessite parfois de projeter le dictionnaire de canal dans cet espace, de sorte à obtenir un ***dictionnaire de réception***. L'extraction du tatouage est couramment effectuée par un démodulateur par corrélation : il s'agit de calculer la corrélation entre le vecteur reçu (et pré-traité) $\hat{\mathbf{v}}$ et les vecteurs du dictionnaire de canal (voire du dictionnaire de réception) supposés connu du récepteur. Le vecteur $\mathbf{d}$ du dictionnaire pour lequel la corrélation est maximum indique finalement le représentant $\hat{m}$ le plus probablement émis. Trois types de corrélation peuvent être envisagés [CMB02] :

- la ***corrélation linéaire*** :

$$c = \frac{1}{N_s} \sum_{i=0}^{N_s-1} \hat{\mathbf{v}}[i]\mathbf{d}[i], \tag{3.3}$$

  où $\hat{\mathbf{v}}[i]$ (respectivement $\mathbf{d}[i]$) désigne la i-ième composante du vecteur $\hat{\mathbf{v}}$ (respectivement $\mathbf{d}$). Ce type de détection est l'équivalent numérique du filtrage adapté utilisé dans les transmissions analogiques [Pro01].

- la ***corrélation normalisée*** :

$$c = \sum_{i=0}^{N_s-1} \frac{\hat{\mathbf{v}}[i]}{||\hat{\mathbf{v}}||} \frac{\mathbf{d}[i]}{||\mathbf{d}||}, \tag{3.4}$$

  où $||\hat{\mathbf{v}}||$ (respectivement $||\mathbf{d}||$) désigne la norme du vecteur $\hat{\mathbf{v}}$ (respectivement $\mathbf{d}$). Ce type de corrélation permet de supprimer la dépendance du récepteur à l'amplitude du signal audio en ne prenant en compte que l'angle formé par le vecteur reçu et les vecteurs du dictionnaire.

- les ***coefficients de corrélation*** :

$$c = \sum_{i=0}^{N_s-1} \frac{\hat{\mathbf{v}}[i] - \mu_{\hat{\mathbf{v}}}}{||\hat{\mathbf{v}} - \mu_{\hat{\mathbf{v}}}||} \frac{\mathbf{d}[i] - \mu_{\mathbf{d}}}{||\mathbf{d} - \mu_{\mathbf{d}}||}, \tag{3.5}$$

  où $\mu_{\hat{\mathbf{v}}}$ (respectivement $\mu_{\mathbf{d}}$) désigne la moyenne du vecteur $\hat{\mathbf{v}}$ (respectivement $\mathbf{d}$). Ce type de corrélation, ne dépendant pas de la moyenne des signaux, permet de rendre le système de tatouage robuste à l'ajout d'offset sur le signal audio.

Des détecteurs non linéaires ont également été proposés tel que JANIS [10] [Fur02, MHSF03] dans un contexte de tatouage sécuritaire. L'objectif est ici de déterminer la présence ou l'absence d'un tatouage. La décision repose alors sur le calcul de la métrique :

$$c = \frac{1}{N_s} \sum_{n_1=0}^{N_s/p-1} \prod_{n_2=0}^{p} \hat{v}[pn_1 + n_2]. \tag{3.6}$$

Cette métrique offre l'avantage de supprimer en moyenne l'influence du signal audio sur le récepteur pour ne conserver que l'information de tatouage.

Dans le cas des fonctions d'insertion substitutives, le signal audio étant supprimé dès la fonction d'insertion, un pré-traitement n'est pas nécessaire. L'extraction du tatouage dépend là encore de la technique d'insertion choisie : si cette dernière consiste à substituer les composantes non perceptibles du signal audio par une modulation BPSK de l'information de tatouage [Gar99],

10. Just Another N-order side-Informed Scheme

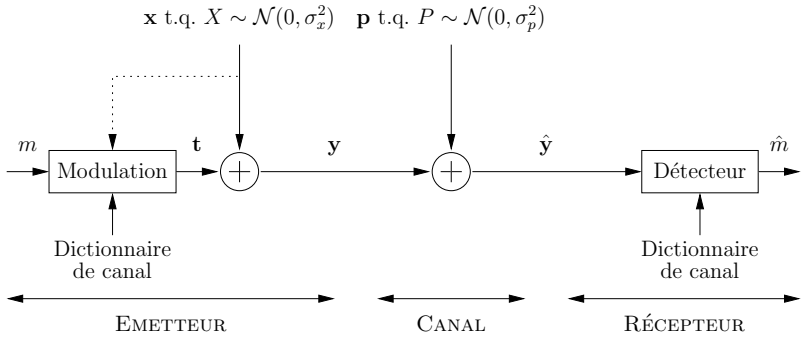


Figure 3.2 – Modèle d'un système de tatouage.

un démodulateur BPSK est utilisé. S'il s'agit d'une quantification du signal hôte [Che00], l'extracteur recherche le point de quantification le plus proche du signal reçu.

Une fois le représentant $\hat{m}$ décidé, il ne reste plus qu'à inverser l'encodage de canal : dans le cas des codes correcteurs d'erreur, le détecteur fait souvent appel à l'algorithme de Viterbi [Pro01].

3.1.2 Performances théoriques

Les performances des systèmes de tatouage sont étroitement liées aux choix du domaine de détection [11] et au caractère informé de la stratégie d'insertion. En effet, le domaine de détection induit la statistique du bruit audio interférant dans le procédé de détection ; la stratégie d'insertion, elle, détermine l'interférence du signal audio sur la détection : dans le cas des fonctions additives non informées, le signal audio est considéré comme un bruit de canal à part entière. Dans le cas des fonctions substitutives ou des fonctions de réjections partielles de l'hôte, son interférence dans la chaîne de transmission est diminuée voire supprimée dès l'étape d'insertion.

Pour évaluer les performances théoriques des systèmes de tatouage, un modèle de la chaîne s'est imposé dans la littérature. Ce modèle projette le système de tatouage directement dans le domaine d'insertion [12]. Le système de tatouage prend alors la forme d'une chaîne de communication à deux bruits additifs présentée figure 3.2. Le premier bruit provient du signal audio lui-même, tandis que le second bruit est issu des perturbations appliquées au signal audio tatoué. Tous deux sont supposés être la réalisation de processus aléatoires indépendants, notés X et P, de lois de probabilité respectives $\mathcal{N}(0, \sigma_x^2)$ et $\mathcal{N}(0, \sigma_p^2)$. L'information à émettre est supposée être un symbole m à valeur dans l'alphabet $\{0, \ldots, M-1\}$. L'encodeur exploite un dictionnaire de canal $\mathcal{D}$ (éventuellement de dimension infinie), structuré sous la forme de M sous-ensembles $\mathcal{D}_m$ chacun associé de manière unique à l'un des symboles possibles. L'encodeur choisit alors le tatouage $\mathbf{t}$ le plus adapté pour représenter m. Ce choix doit satisfaire la contrainte d'inaudibilité propre aux systèmes de tatouage. Cette contrainte est ici modélisée sous la forme d'une contrainte de puissance : la puissance du tatouage $\mathbf{t}$, notée σ_t^2, doit être choisie inférieure ou égale à une valeur seuil, notée β :

$$\sigma_t^2 \leq \beta. \tag{3.7}$$

Le détecteur recherche le vecteur du dictionnaire le plus proche du signal reçu $\hat{\mathbf{y}}$. Ce vecteur

11. c'est à dire le domaine d'insertion ayant éventuellement subi les pré-traitements du récepteur.
12. supposant que le domaine d'insertion et le domaine de détection sont identiques et supprimant ainsi les transformations $\mathcal{T}$ à l'émetteur et au récepteur

appartenant à l'un des sous-ensembles $\mathcal{D}_m$, le symbole reçu est alors m.

Dans le cas des systèmes non informés, c'est à dire n'exploitant pas la connaissance du signal audio $\mathbf{x}$ à l'émetteur, le choix du tatouage $\mathbf{t}$ est indépendant de $\mathbf{x}$. Le dictionnaire de canal contient autant d'éléments que de symboles possibles. Il n'existe donc qu'un seul tatouage représentant le symbole m. Le signal audio $\mathbf{x}$ est alors vu comme un bruit venant perturber la détection, au même titre que le bruit de perturbation $\mathbf{p}$. La chaîne de tatouage est finalement une chaîne de communication standard de bruit BBAG, réalisation du processus aléatoire $X+P$. Ces performances en terme de capacité d'insertion C_{ni} sont données par [Sha48] :

$$C_{ni} = \frac{1}{2} log_2 \left(1 + \frac{\sigma_t^2}{\sigma_x^2 + \sigma_p^2} \right). \tag{3.8}$$

Elles sont donc inévitablement limitées par l'influence du signal audio sur le procédé de détection.

Dans le cas des systèmes informés (également appelés canaux avec information de bord), les travaux de Costa [Cos83] dans la lignée des contributions de Shannon [Sha58], de Gel'fand et Pinsker [GP80] ont montré que la capacité de transmission pouvait être nettement améliorée dès lors que la connaissance *a priori* du signal audio $\mathbf{x}$ à l'émetteur est exploitée. Elle devient :

$$C_i = \frac{1}{2} log_2 \left(1 + \frac{\sigma_t^2}{\sigma_p^2} \right). \tag{3.9}$$

Cette capacité est celle d'un canal BBAG soumis uniquement au bruit de perturbation $\mathbf{p}$. Il est donc possible de construire un système qui s'affranchisse complètement du bruit audio. Ce gain, en terme de capacité, laisse entrevoir des systèmes de tatouage audio plus performants (en terme de fiabilité de transmission et de robustesse aux perturbations de canal), à condition de choisir une stratégie d'insertion adaptée exploitant la connaissance *a priori* du signal audio.

3.2 Spécificités des systèmes non informés

De nombreuses contributions [Kim03, BBC+04] proposent un état de l'art des systèmes de tatouage audio non informés. Les éléments de base de leur conception ayant été évoqués précédemment, nous ne reviendrons donc pas sur la description détaillée des schémas de tatouage audio non informé. Nous préférons consacrer cette courte section à mettre en évidence les axes de recherches novateurs qui se sont développés parallèlement aux systèmes de tatouage informés et à fournir un ordre de grandeur des performances de ces systèmes.

3.2.1 Axes de recherche

Les travaux ayant trait aux systèmes non informés se sont principalement portés vers la recherche de récepteurs efficaces, dont les objectifs sont de diminuer l'influence du signal audio lors du procédé de détection. Nous nous sommes plus particulièrement intéressé à deux d'entre eux : le système de Larbi [Lar05] et du système de Cvejic *et al.* [CTS03].

Larbi dans sa thèse [Lar05] étudie l'apport des structures d'égalisation sur les performances d'un système de tatouage. Elle propose un schéma de réception particulièrement intéressant basé sur l'égaliseur de Wiener. Le schéma de son système est présenté figure 3.3 ; il est construit dans la lignée du système proposé par Gomes [Gom02]. L'information à insérer est modulée à l'aide d'un dictionnaire contenant des vecteurs étalés spectralement. Le signal modulé résultant $v(n)$

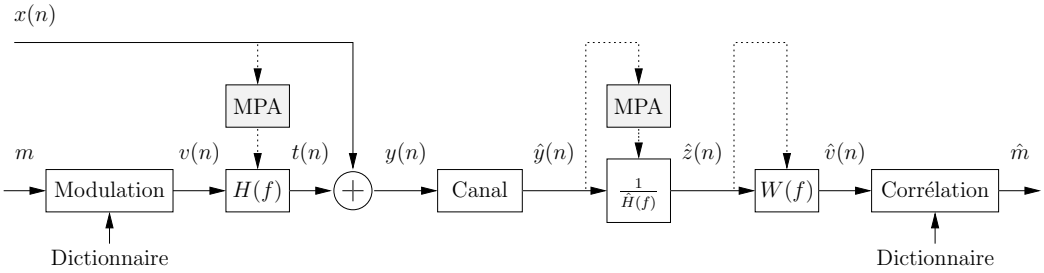


Figure 3.3 – Système de tatouage non informé de Larbi [Lar05].

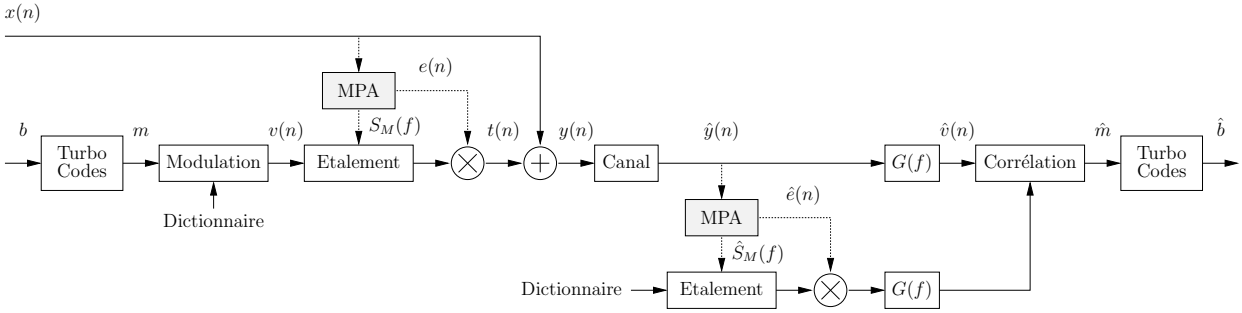


Figure 3.4 – Système de tatouage non informé de Cvejic *et al.* [CTS03].

est ensuite mis en forme perceptuellement en exploitant un Modèle PsychoAcoustique (MPA). Cette mise en forme s'interprète sous la forme d'une opération de filtrage par un filtre $H(f)$. Le signal de tatouage $t(n)$ ainsi obtenu est ensuite ajouté au signal audio $x(n)$. La détection exploite la procédure d'égalisation suivante : un filtre zero-forcing $\frac{1}{\hat{H}(f)}$ est utilisé pour inverser la mise en forme spectrale du signal modulé par $H(f)$. $\hat{H}(f)$ est ici une estimation du filtre $H(f)$, calculée à partir du signal audio tatoué $\hat{y}(n)$ (éventuellement détérioré par le canal). A supposer $\hat{H}(f) = H(f)$, le signal ainsi obtenu $\hat{z}(n)$ peut s'exprimer sous la forme $\hat{z}(n) = v(n) + r(n)$ où $r(n)$, résultat du filtrage du signal audio $x(n)$ par $\frac{1}{\hat{H}(f)}$, se révèle être une version blanchie du signal audio [13]. Dans la pratique, $\hat{H}(f) \approx H(f)$ et $\hat{z}(n) \approx v(n) + r(n)$; de plus, le rapport de puissance entre le signal modulé recherché et le bruit audio filtré $r(n)$ malgré l'effet blanchissant de $\frac{1}{\hat{H}(f)}$ est peu favorable à la détection de $v(n)$. Un filtre de Wiener $W(f)$ non causal est donc synthétisé dans l'objectif d'estimer le signal modulé $v(n)$ à partir de l'observation $\hat{z}(n)$. Les coefficients de ce filtre sont calculés de sorte à minimiser l'erreur quadratique moyenne $E\left[||\hat{v}(n) - v(n)||^2\right]$ entre le signal $\hat{v}(n)$ obtenu en sortie du filtre de Wiener et le signal modulé $v(n)$. Ils sont, sous forme vectorielle :

$$\mathbf{w} = R_{\hat{\mathbf{z}}}^{-1} r_{\mathbf{v}}, \tag{3.10}$$

où $R_{\hat{\mathbf{z}}}$ est la matrice d'autocovariance du signal $\hat{z}(n)$ et $r_{\mathbf{v}}$ est la fonction d'autocovariance

13. En effet, le MPA choisi pour ce système confère au filtre de mise en forme $H(f)$ une allure proche de la Densité Spectrale de Puissance (DSP) du signal audio. Ainsi, filtrer le signal audio par $\frac{1}{\hat{H}(f)}$ revient à inverser la DSP de ce signal donc à le blanchir.

estimée du signal $v(n)$. Le détecteur est un démodulateur par corrélation qui utilise le même dictionnaire que celui d'émission. Ce schéma de tatouage, de faible complexité, permet une mise en œuvre temps-réel du système.

Le système proposé par Cvejic *et al.* est également un système de tatouage par étalement de spectre dont l'insertion s'effectue de manière additive dans le domaine temporel. Son schéma de principe est présenté figure 3.4. L'information à insérer est encodé par un turbo code. Elle est là encore modulée à l'aide d'un dictionnaire contenant des vecteurs étalés spectralement. Le signal modulé $v(n)$ qui en résulte est mis en forme perceptuellement en exploitant un MPA. Cette mise en forme est effectuée dans un premier temps dans le domaine fréquentiel en multipliant le spectre du signal modulé par le seuil de masquage, puis en modulant l'amplitude du signal mis en forme par l'enveloppe $e(n)$ du signal audio $x(n)$. De cette mise en forme résulte le signal de tatouage $t(n)$ qui est ensuite ajouté au signal audio $x(n)$. La procédure de détection exploite un filtrage blanchissant basé sur des filtres de Savitzky-Golay [14]. La décision repose sur un démodulateur par corrélation, qui compare le signal filtré $\hat{v}(n)$ au vecteur d'un dictionnaire de réception ; ce dictionnaire contient une estimation des vecteurs du dictionnaire de canal mis en forme spectralement et blanchis.

3.2.2 Performances

Comparer les performances des systèmes de tatouage en terme de fiabilité de transmission et de robustesse face aux perturbations s'avère relativement ardu, et ce pour deux raisons.

La première tient au fait que la majorité de ces systèmes est dédiée à la protection de la propriété intellectuelle. L'évaluation des performances du système est donc établie pour des critères différents de ceux que nous avons considérés dans le protocole expérimental, décrit section 2.4. La mesure du TEB est remplacée par la probabilité d'erreur de détection et de fausse alarme. Un système dédié à la protection cherche en effet à établir la présence ou l'absence de tatouage. Le tatouage doit donc être détecté avec le plus de certitude possible s'il est présent, et ne pas être détecté s'il est absent. Le débit de transmission est volontairement choisi faible pour garantir la robustesse aux perturbations. Les débits considérés sont majoritairement inférieurs à 50 bit/s. Dans certains systèmes, ils sont même dépendants du signal audio : dans [MGN02], le tatouage substitue les composantes non significatives du signal audio (inférieures à un seuil de masquage obtenu après analyse psychoacoustique du signal audio). Le débit de transmission dépend alors du nombre de composantes du signal audio inférieures au seuil de masquage.

La seconde raison tient à la contrainte d'inaudibilité. Les systèmes de tatouage sont en effet conçus pour établir un compromis entre inaudibilité, fiabilité de transmission et robustesse. Pour comparer équitablement leurs performances, il faudrait pouvoir évaluer les TEBs à "inaudibilité équivalente", c'est à dire pour un signal de tatouage induisant la même dégradation auditive lorsqu'il est inséré au signal audio. L'utilisation de l'outil de mesure objective de la dégradation sonore PEAQ (proposé précédemment) n'étant pas encore répandu, nous ne disposons que peu d'informations sur les performances des systèmes en matière d'inaudibilité. Ces remarques étant faites, nous nous proposons de donner, à titre indicatif, des ordres de grandeur des performances des systèmes de la littérature.

A notre connaissance, seuls les deux systèmes présentés précédemment, le système de Cvejic *et al.* [CTS03] et celui de Larbi [Lar05], proposent une évaluation des performances sur des signaux audio réels pour notre contexte de transmission de l'information. La figure 3.5 présente les TEBs

14. Ces filtres sont spécifiquement utilisés pour lisser des signaux dont le spectre est à large bande.

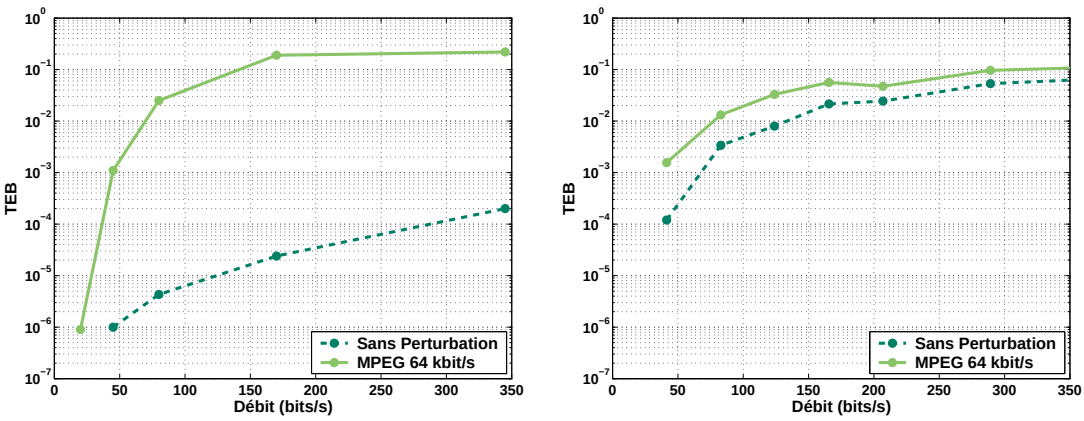


Figure 3.5 – Performances des deux systèmes de tatouage additif par étalement de spectre dans le domaine temporel : (a) le système de Cvejic *et al.* et (b) le système de Larbi [Lar05], chacun dans le cas d'un canal sans perturbation et avec compression MPEG à 64 kbit/s.

fournis par les deux auteurs, dans le cas d'un canal sans perturbation puis avec compression MPEG à 64 kbit/s. Le système de Cvejic semble très efficace lorsque le canal ne subit pas de perturbations, mais ses performances sont très fortement dégradées à débit de transmission élevé dès lors qu'une compression MPEG est entreprise. Le système de Larbi est moins performant que le système de Cvejic mais plus robuste à la compression MPEG.

3.3 Spécificités des systèmes de tatouage informés

Les recherches ayant trait à la construction de stratégies d'insertion informée ont débuté lorsque Cox *et al.* [CMM99] établissent l'analogie entre les systèmes de tatouage et les chaînes de communication avec information adjacente introduite par Shannon [Sha58]. Les travaux de Costa [Cos83] sur les canaux avec information de bord montrent que les performances du système en terme de capacité peuvent être rendues indépendantes du signal audio à condition d'exploiter la connaissance *a priori* du signal audio à l'émetteur. Il propose également un encodage canal, depuis appelé ***schéma de Costa***, qui permet d'atteindre la capacité théorique de la chaîne avec information adjacente. Ce schéma, décrit par la suite, reste théorique et loin des réalités imposées par la mise en œuvre effective et efficace d'un système de tatouage. De nombreux systèmes de tatouage informés en ont été déduits : la fonction d'insertion est modifiée pour choisir un tatouage $\mathbf{t}$ satisfaisant la robustesse de la transmission à un bruit de canal supposé additif et l'inaudibilité du tatouage. Ces stratégies s'attachent aussi bien au tatouage des images, de la vidéo que de l'audio. Dans les sections à venir, nous nous intéresserons spécifiquement aux stratégies proposées dans le domaine de l'audio et aux stratégies génériques (c'est à dire définies pour un support numérique quelconque).

3.3.1 Schéma de Costa

Le schéma de Costa s'attache au modèle théorique de la chaîne de tatouage présenté figure 3.2.

Une chaîne de communication standard, n'ayant pas connaissance du bruit audio, à l'émetteur, chercherait à diminuer l'influence du bruit audio lors du procédé de détection. Une chaîne de communication avec information de bord en ferait autant si la puissance du bruit audio est plus faible que celle du tatouage : en effet, dans ce contexte, une partie de l'énergie du tatouage pourrait être utilisée pour annuler le bruit tandis que l'autre servirait à transmettre l'information. Costa, lui, s'intéresse au cas où la puissance du tatouage est nettement inférieure au bruit audio. Puisque ce rapport de puissance ne permet pas d'annuler les effets du bruit audio sur la transmission, l'encodeur doit au contraire s'adapter au canal. Pour ce faire, il doit choisir un vecteur du dictionnaire de canal porteur de l'information à émettre et dans la direction du signal audio. Il profite ainsi du signal audio pour transmettre efficacement l'information.

3.3.1.1 Principe

Plus concrètement, Costa propose de partitioner le dictionnaire de canal en M sous-dictionnaires $\mathcal{D}_m$. Chaque sous-dictionnaire contient un ensemble (de grande taille) de vecteurs tous susceptibles de moduler l'information m. Lors de l'insertion de l'information m, l'encodeur choisit le vecteur $\mathbf{d}$ du sous-dictionnaire $\mathcal{D}_m$ associé à m tel que D et le signal audio $\mathbf{x}$ soient typiquement joints [15]. Le tatouage transmis est alors donné par :

$$\mathbf{t} = \mathbf{d} - \alpha\mathbf{x}, \tag{3.11}$$

de sorte que le signal audio tatoué s'exprime sous la forme :

$$\mathbf{y} = \mathbf{x} + \mathbf{t} = (1-\alpha)\,\mathbf{x} + \mathbf{d}. \tag{3.12}$$

La détection s'attache alors à rechercher parmi tous les vecteurs du dictionnaire le vecteur $\hat{\mathbf{d}}$ tel que $\hat{\mathbf{d}}$ et $\mathbf{y}$ soient typiquement joints. $\hat{\mathbf{d}}$ appartenant à un unique sous-dictionnaire $\mathcal{D}_{\hat{m}}$, l'information reçue est $\hat{m}$.

Costa démontre que la capacité d'insertion est maximale et devient $C_i = \frac{1}{2}\log_2\left(1+\frac{\sigma_t^2}{\sigma_p^2}\right)$ à condition de choisir :

$$\alpha = \frac{\sigma_t^2}{\sigma_t^2+\sigma_p^2}. \tag{3.13}$$

La stratégie d'insertion optimale prend donc la forme d'une réjection de l'hôte. Comme le fait remarquer Pateux [BBC+04], cette réjection n'est que partielle (puisque $\alpha < 1$). Le choix du vecteur du dictionnaire doit permettre d'amener le signal tatoué vers la région de détection de l'information m à émettre.

3.3.1.2 Limites

Bien que le schéma de Costa résolve le problème du tatouage pour un modèle théorique, sa mise en œuvre pratique révèle plusieurs problèmes [BPG02] :

- la mesure de la distorsion est inadéquate pour les systèmes de tatouage. Costa établit la contrainte d'inaudibilité par l'EQM entre le signal audio tatoué $\mathbf{y}$ et le signal audio original $\mathbf{x}$:

$$\mathrm{EQM} = E\left[||\mathbf{y}-\mathbf{x}||^2\right] = E\left[||\mathbf{t}||^2\right]. \tag{3.14}$$

15. Les deux vecteurs $\mathbf{d}$ et $\mathbf{x}$ sont typiquement joints s'il existe un coefficient α tel que $E\left[(\mathbf{d}-\alpha\mathbf{x})\mathbf{x}\right] = 0$, où $E[X]$ désigne l'espérance mathématique de X.

Cette mesure, comme nous l'avons précisé précédemment, ne prend pas en compte les distorsions locales introduites par l'insertion du tatouage. En effet, garantir une EQM constante permet de compenser globalement des augmentations locales de la puissance du tatouage, qui peuvent malheureusement induire des dégradations perceptuelles significatives.

- la construction du dictionnaire de canal est difficile. Ce dictionnaire doit être de grande taille, pour garantir une bonne répartition des vecteurs dans l'espace d'insertion : l'existence d'un vecteur $\mathbf{d}$ porteur de l'information m à proximité du signal audio $\mathbf{x}$ est ainsi garantie. La dimension élevée du dictionnaire rend irréaliste la mise en œuvre pratique puisque le temps de calcul nécessaire à la recherche du vecteur (à l'émetteur comme au récepteur) $\mathbf{d}$ est très élevé.

- la modélisation des deux bruits de la chaîne, le bruit audio et le bruit de perturbation, sous la forme d'un processus aléatoire blanc Gaussien est irréaliste. Costa ne spécifie pas le domaine d'insertion dans lequel le signal audio et le bruit de perturbation peuvent être supposés BBAG. Or, comme nous l'avons spécifié précédemment, le choix d'un tel domaine est encore incertain. De plus, en pratique, les deux bruits sont rarement Gaussiens [CMB02]. Et l'indépendance entre les deux bruits ne peut être garantie.

- la stratégie d'insertion suppose la statistique du bruit de perturbation connue, ce qui n'est pas le cas en pratique lors de la construction du signal audio tatoué.

Conclure que la capacité d'un système de tatouage pratique est effectivement indépendante du signal audio s'avère prématuré [CMB02]. Pour autant, les résultats de Costa laissent présager un gain non négligeable à construire des systèmes de tatouage avec prise en compte du bruit audio à l'émetteur. De nombreux systèmes adaptant le schéma de Costa à un modèle de tatouage plus réaliste ont donc été proposés. Les premiers s'intéressent au cas des fonctions d'insertion additives basées sur un récepteur par corrélation ; les seconds choisissent une fonction d'insertion substitutive par quantification. Ces deux catégories font l'objet des sections à venir.

3.3.2 Fonctions d'insertion additive informée

3.3.2.1 Vers un problème d'optimisation

La fonction d'insertion informée doit concilier au mieux l'inaudibilité du tatouage et sa détection robuste. Ces performances sont donc intrinsèquement liées au choix du détecteur. Miller *et al.* dans [MCB00] s'intéresse plus particulièrement aux détecteurs par corrélation et propose d'établir avec précision les alternatives en matière de fonction d'insertion robuste. Dans cet objectif, il propose une interprétation géométrique de la stratégie d'insertion. Deux régions d'intérêt se dégagent de cette interprétation :

- la ***région de distorsion acceptable***. Cette région est l'ensemble des signaux audio tatoués $\mathbf{y}$ perceptuellement proches du signal audio original $\mathbf{x}$,

- la ***région de détection*** qui regroupe l'ensemble des signaux audio tatoués pour lesquels l'information à émettre m est correctement détectée. Cette région peut être restreinte à un sous-ensemble appelé ***région de détection robuste*** pour laquelle la détection de l'information k est robuste à un bruit de canal.

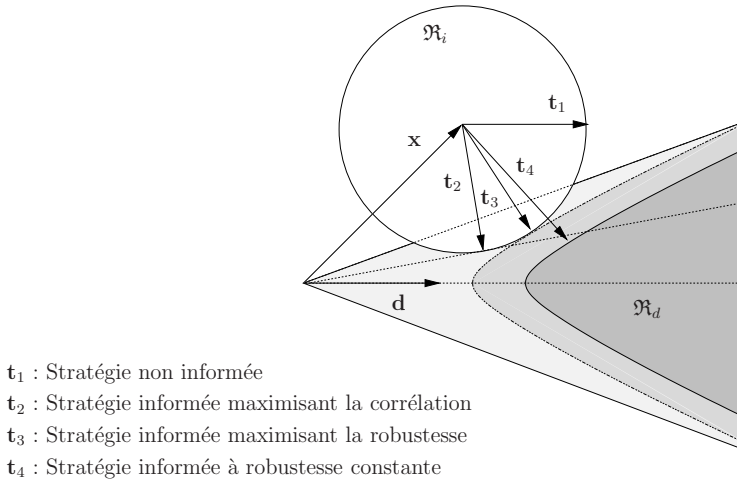


Figure 3.6 – Interprétation géométrique de la stratégie d'insertion proposée par Miller *et al.* (extrait de [MDC02]).

Dans le but de visualiser ces deux régions, les auteurs reprennent le modèle du schéma de Costa présenté section 3.3.1. L'information à émettre est supposée être inaudible si elle est encodée par un tatouage $\mathbf{t}$ de puissance limitée $\sigma_t^2 \leq \beta$. La région d'inaudibilité est donc une hyper-sphère centrée autour du signal audio $\mathbf{x}$ et de rayon σ_t. Cette information est également supposée détectée si la corrélation entre le signal audio tatoué bruité $\mathbf{y} = \mathbf{x} + \mathbf{t} + \mathbf{p}$ et le vecteur $\mathbf{d}$ du dictionnaire de canal est maximal (au regard des corrélations avec les autres vecteurs du dictionnaire). En l'absence de bruit, la région de détection est donc un hyper-cône centré autour du vecteur $\mathbf{d}$; en présence de bruit, cet hypercône s'éloigne progressivement de l'origine (au fur et à mesure que le bruit augmente). La figure 3.6 visualise ces régions d'intérêt dans différentes configurations : la région d'inaudibilité est le disque $\mathfrak{R}_i$, la région de détection robuste est l'hypercône $\mathfrak{R}_d$, devenant de plus en plus grisé à mesure que le bruit de perturbation devient de plus en plus important. Les auteurs établissent que concilier l'inaudibilité et la détection robuste nécessite donc de choisir un tatouage $\mathbf{t}$ à l'intersection de ces deux régions.

Un système non informé n'exploitant pas la connaissance a priori du signal audio conduirait à choisir pour tatouage le vecteur $\mathbf{d}$. Ce choix est représenté par le tatouage $\mathbf{t}_1$. Dans un contexte informé, les auteurs établissent trois stratégies d'insertion possibles, conduisant à trois signaux de tatouage différents :

- ***la stratégie maximisant la corrélation***. La détection exploitant la corrélation entre le signal reçu $\hat{\mathbf{y}} = \mathbf{x} + \mathbf{t} + \mathbf{p}$ et le vecteur $\mathbf{d}$, le tatouage $\mathbf{t}$ peut être choisi de sorte que $\hat{\mathbf{y}}$ soit dans la région de détection correcte (en l'absence de bruit) et dans la région d'inaudibilité, pour une valeur de la corrélation maximale. Le tatouage choisi est alors le vecteur $\mathbf{t}_2$. Cette stratégie, bien qu'étant théoriquement la plus efficace, ne peut être mise en pratique : en effet, le bruit de perturbation $\mathbf{p}$ n'est pas connu lors de la phase d'émission.

- ***la stratégie maximisant la robustesse***, c'est à dire choisissant le tatouage inaudible garantissant la détection correcte pour un bruit de perturbation maximal. Cette stratégie nécessite de choisir un critère modélisant la robustesse aux bruits de perturbations. Dans [MCB00], Miller *et al.* suggère de remplacer le vecteur de bruit $\mathbf{p}$ par un paramètre de robustesse à maximiser. Cette maximisation conduit alors à choisir pour tatouage $\mathbf{t}$

l'unique point de tangence entre la région d'inaudibilité et un hypercône $\mathfrak{R}'_d$ caractérisant la robustesse de la transmission à un bruit donné. Ce bruit est le bruit maximum admissible n'entraînant pas d'erreur de détection. Cette stratégie donne lieu au tatouage $\mathbf{t}_3$.

- ***la stratégie garantissant une robustesse constante aux perturbations***. Pour un bruit de perturbation donné, le tatouage est choisi pour garantir la détection correcte de l'information émise au détriment (parfois) de la contrainte d'inaudibilité. Plusieurs solutions existant pour un paramètre de robustesse donné, l'algorithme choisit nécessairement celui dont la distorsion perceptuelle est la plus petite. En effet, le tatouage solution $\mathbf{t}_4$ se trouve à l'extérieur de l'hyper-sphère d'inaudibilité. Cette stratégie d'insertion peut paraître obsolète puisque ne respectant pas la contrainte première d'un système de tatouage c'est à dire la transparence de la marque. Pourtant, comme nous allons le voir par la suite, cette stratégie est la plus répandue parmi les systèmes de tatouage audio informés proposés dans la littérature.

Au regard de ces trois stratégies, établir en pratique une fonction d'insertion adaptée conduit à un problème d'optimisation. Ce problème peut s'exprimer de deux façons [CMB02] :

(1) trouver le tatouage qui maximise la robustesse du système aux perturbations sous la contrainte de limiter la distorsion perceptuelle.

(2) trouver le tatouage qui minimise la distorsion perceptuelle sous la contrainte de garantir la robustesse du système à une perturbation donnée.

Les systèmes de tatouage avec fonctions d'insertion informée additives s'inscrivent dans ces deux catégories.

Minimisation de la distorsion à robustesse fixée

La majorité des systèmes informés trouvés dans la littérature propose une mise en pratique de la stratégie de robustesse aux perturbations constantes de Miller.

Lobo Guerrero [LGP04] s'intéresse au cas de l'insertion informée d'un message binaire. Le système proposé est un système additif dans le domaine temporel : l'insertion du bit b débute par une modulation binaire antipodale exploitant un vecteur étalé spectralement $\mathbf{d}$ et une mise en forme spectrale (par pondération du spectre par un seuil de masquage calculé à l'aide du modèle psychoacoustique n°1 de MPEG), donnant lieu au signal $\mathbf{v} = (2b-1)\tilde{\mathbf{d}}$, où $\tilde{\mathbf{d}}$ est le vecteur de modulation mis en forme spectralement. Le tatouage, obtenu par pondération du signal $\mathbf{v}$ par un facteur d'amplitude α :

$$\mathbf{t} = \alpha\mathbf{v} = \alpha\,(2b-1)\,\tilde{\mathbf{d}} \tag{3.15}$$

est ensuite ajouté au signal audio $\mathbf{x}$. La détection est effectuée par une étape de blanchiment suivie d'un démodulateur par corrélation [16] qui compare le signal reçu à une version filtrée des vecteurs du dictionnaire de canal. La prise en compte du signal audio est effectuée par un module de contrôle de la puissance du tatouage : ce module choisit le facteur d'amplitude α qui garantit la détection robuste du bit b. Pour ce faire, il simule la procédure de détection et modifie la valeur de α jusqu'à ce que la corrélation entre le signal audio tatoué $\mathbf{y} = \mathbf{x} + \mathbf{t} = \mathbf{x} + \alpha\mathbf{v}$ et le vecteur du dictionnaire de canal à détecter atteigne une valeur seuil σ_r caractérisant la robustesse du système au bruit.

Dans [MS04], Mayer et Silva généralisent cette même stratégie aux modulations CDMA permettant l'insertion simultanée de plusieurs bits d'information $\{b_m\}_{m\in[\![0,M-1]\!]}$. Elle exploite un

16. LoboGuerrero opte en réalité pour un filtrage adapté, équivalent analogique du démodulateur par corrélation.

dictionnaire de M vecteurs orthogonaux étalés spectralement notés $\{d_m\}_{m\in[\![0,M-1]\!]}$, conduisant au signal de tatouage :

$$\mathbf{t} = \sum_{m=0}^{M-1} \alpha_m \left(2b_m - 1\right) \tilde{\mathbf{d}}_m, \tag{3.16}$$

où $\tilde{\mathbf{d}}_m$ désigne la mise en forme perceptuelle des vecteurs du dictionnaire. La détection est simulée jusqu'à ce que les métriques de détection atteignent une valeur seuil σ_r caractérisant la robustesse du système. Dans les deux cas, le paramètre de robustesse au bruit est fixé en fonction de l'application. La transparence du tatouage n'est contrôlée qu'après insertion du message ; elle n'est donc pas garantie lors de l'insertion du tatouage.

Le cas des modulations M-aire est traité par Döerr [Doë01, LDC04] : le système est toujours un système additif par étalement de spectre. Le codage de source associé à chaque K-uplet binaire un symbole m parmi $\{0, \ldots, M-1\}$ avec $M = 2^K$. La modulation exploite alors un dictionnaire de M vecteurs orthogonaux, de sorte que le symbole m est détecté si la corrélation entre le signal reçu $\mathbf{y} = \mathbf{x} + \mathbf{t}$ et le vecteur $\mathbf{d}_m$ est maximale (au regard des corrélations avec les autres vecteurs). Döerr propose un algorithme d'insertion itératif permettant de construire pas à pas le tatouage jusqu'à atteindre les conditions de transmission robuste. Ces conditions sont évaluées par le démodulateur par corrélation : pour transmettre sans erreur l'information k, la corrélation entre le signal audio tatoué détérioré $\hat{\mathbf{y}}$ et le vecteur $\mathbf{d}_k$ doit être supérieure aux corrélations avec les autres vecteurs du dictionnaire de canal, quel que soit le bruit de perturbation $\mathbf{p}$.

$$\forall m \neq k, (\mathbf{x}+\mathbf{t}+\mathbf{p})^t \mathbf{d}_k \geq (\mathbf{x}+\mathbf{t}+\mathbf{p})^t \mathbf{d}_m \Leftrightarrow (\mathbf{x}+\mathbf{t})^t \frac{\mathbf{d}_k - \mathbf{d}_m}{||\mathbf{d}_k - \mathbf{d}_m||} \geq -\mathbf{p}^t \frac{\mathbf{d}_k - \mathbf{d}_m}{||\mathbf{d}_k - \mathbf{d}_m||}. \tag{3.17}$$

Doërr en déduit que la transmission sera d'autant plus robuste que l'écart entre $(\mathbf{x}+\mathbf{t})^t \mathbf{d}_k$ et les $(\mathbf{x}+\mathbf{t})^t \mathbf{d}_m$ est grand. Sa stratégie consiste donc à rechercher le tatouage $\mathbf{t}$ qui maximise la métrique :

$$R(\mathbf{t}) = \min_{m\neq k} (\mathbf{x}+\mathbf{t})^t \left(\frac{\mathbf{d}_k - \mathbf{d}_m}{||\mathbf{d}_k - \mathbf{d}_m||} \right). \tag{3.18}$$

L'indice m pour lequel $R(\mathbf{t})$ atteint sa valeur minimum est l'indice du vecteur le plus compétitif, c'est à dire celui ayant le plus de risque d'être détecté à la place du vecteur $\mathbf{d}_k$. Pour maximiser la robustesse de son système aux perturbations, Döerr se fixe une valeur seuil σ_r^2 de $R(\mathbf{t})$ à atteindre. L'algorithme consiste alors à construire le tatouage en lui ajoutant étape par étape une contribution dans la direction $\frac{\mathbf{d}_k - \mathbf{d}_m}{||\mathbf{d}_k - \mathbf{d}_m||}$, où m est l'indice du vecteur le plus compétitif. L'algorithme s'arrête dès que la valeur seuil de $R(\mathbf{t})$ est atteinte. Cet algorithme, bien qu'efficace, néglige malheureusement la contrainte d'imperceptibilité. En effet, la distorsion introduite par le tatouage n'est contrôlée, là encore, qu'après insertion de l'information et n'est donc pas garanti pendant l'insertion du tatouage. D'autre part, plusieurs tatouages peuvent être solutions de l'algorithme (puisque l'algorithme s'arrête dès lors qu'un tatouage permet d'atteindre la robustesse escomptée) ; on ne peut donc garantir que le tatouage obtenu à l'issu de l'algorithme soit celui qui minimise la distorsion perceptuelle.

Maximisation de la robustesse à distorsion perceptuelle limitée

Malvar et Florencio [MF03] sont les seuls, à notre connaissance, à proposer une stratégie d'insertion, l'ISS [17], ayant pour contrainte de limiter la distorsion perceptuelle qu'il introduit.

17. Improved Spread-Spectrum

Ils s'intéressent à un modèle de la chaîne de tatouage (proche de celui de Costa) dans lequel : la construction du tatouage $\mathbf{t}$ est basée sur une modulation binaire antipodale du bit à émettre (cette modulation exploite un vecteur $\mathbf{d}$ étalé spectralement) ; l'insertion est additive ; la distorsion auditive est contrôlée par un critère d'EQM : cette EQM, correspondant à $E\left[||\mathbf{y}-\mathbf{x}||^2\right] = E\left[||\mathbf{t}||^2\right]$, est maintenue à une valeur limite σ_r^2. Plutôt que de choisir pour tatouage $\mathbf{t} = (2b-1)\,\mathbf{d}$ (comme ce serait le cas dans une stratégie non informée), les auteurs proposent un tatouage proche de l'expression du modèle de Costa :

$$\mathbf{t} = (\alpha\,(2b-1) - \lambda x)\,\mathbf{d}, \tag{3.19}$$

où $x = \dfrac{\mathbf{x}^t\mathbf{d}}{||\mathbf{d}||^2}$ est la projection du signal audio dans la direction de détection $\mathbf{d}$. L'expression du tatouage devient similaire à celle proposée par Costa : il annule une partie de l'interférence du signal audio sur le procédé de détection tout en transmettant l'information b. Les paramètres α et β sont choisis de sorte à conserver la distorsion perceptuelle constante ($E\left[||\mathbf{t}||^2\right] = \sigma_r^2$) et à minimiser la probabilité d'erreur de transmission, lorsque le système est soumis à une perturbation de canal de type BBAG $\mathcal{N}(0,\sigma_p^2)$. La maximisation de la robustesse est donc obtenue à condition de connaître la puissance du bruit σ_p^2. Bien que les performances théoriques de son système soient très prometteuses, sa mise en œuvre pratique révèle deux problèmes. D'une part, mesurer la distorsion perceptuelle par un critère EQM ne permet pas de contrôler la distorsion locale introduite par le tatouage. Malvar et Florencio eux-même le reconnaissent et ajoutent une étape limitant localement la puissance du tatouage : cette étape consiste à ne pas tatouer les fenêtres où la distorsion/la puissance du tatouage serait trop importante c'est à dire supérieure à une valeur seuil K. Le choix de cette valeur seuil n'est malheureusement pas précisé. D'autre part, les paramètres d'insertion α et λ dépendent de la puissance du bruit de canal, alors que ce bruit n'est pas connu lors de la phase d'émission.

Cvejic et Sapponën [CS03] proposent une application de la stratégie de Malvar et Florencio à un système de type additif par étalement de spectre avec mise en forme perceptuelle de l'information à tatouer par un modèle psychoacoustique. La pattern insérée est mise en forme spectralement pour que son spectre ait l'allure du seuil de masquage. Sa puissance est modifiée en suivant l'algorithme de Malvar et Florencio ; de fait, l'inaudibilité locale de la pattern malgré la mise en forme perceptuelle n'est pas garantie.

Ce travail de thèse s'inscrit dans ce contexte encore peu étudié ; notre objectif est de proposer une stratégie d'insertion informée qui maximise la robustesse du système aux distorsions et limite la distorsion perceptuelle - cette fois - localement.

3.3.2.2 Vers un dictionnaire structuré

Miller [Mil01] propose de modifier un système de tatouage additif par étalement de spectre de sorte à approcher la structure proposée par Costa. Le dictionnaire de canal est alors structuré sous la forme de M sous-dictionnaires. Chaque sous-dictionnaire $\mathcal{D}_m$ contient un ensemble de Q vecteurs, tous représentant la même information m. Le détecteur par corrélation est simulé à l'émetteur pour rechercher parmi ces vecteurs celui ayant le plus de chance d'être détecté, c'est à dire celui dont la corrélation avec le signal audio tatoué est la plus grande. Miller met en évidence l'efficacité de sa méthode en comparant le rapport de puissance entre le tatouage et le bruit audio (RSB) : plus le nombre de vecteurs par sous-dictionnaire augmente, meilleur est le RSB, et donc vraisemblablement, meilleure est la détection. En contre-partie, le détecteur nécessitant une recherche exhaustive du vecteur de $\mathcal{D}$ de corrélation maximale avec le signal

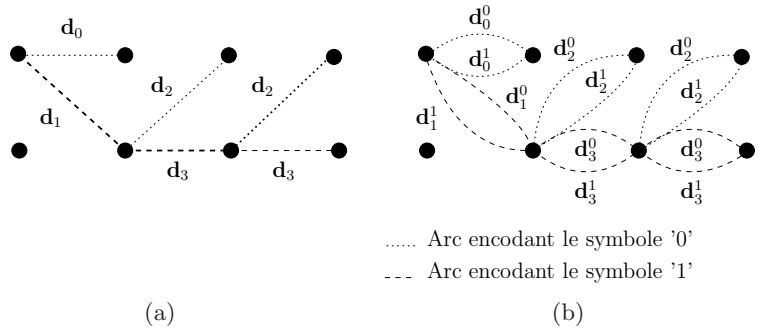


Figure 3.7 – Modulation en treillis proposée par Ungerboëck [Ung87] et reprise par Miller [MDC02] pour les besoins du tatouage dans le cas non informé (a) et dans le cas informé (b).

reçu, le coût en temps de calcul devient prohibitif dès que Q devient trop grand. Construire un dictionnaire structuré facilitant la recherche du vecteur reçu devient une nécessité.

Dans cet objectif, Miller propose de réaliser l'insertion de la séquence de symboles en utilisant un treillis informé [MDC02]. Cette technique a été introduite par Ungerboëck [Ung87] puis largement utilisée pour les communications numériques. Des usages spécifiques ont aussi été réalisés par exemple pour le codage des signaux de parole [Pop95]. Dans un treillis conventionnel, tel celui présenté figure 3.7 (a), l'encodage d'une suite de L symboles emprunte un unique chemin (suite de L arcs) : à un nœud donné du treillis, le symbole m est encodé en choisissant l'unique arc parmi les M possibles qui le représente. Cet arc est associé à un vecteur du dictionnaire qui est utilisé pour la modulation. A titre d'exemple, la suite de symboles $'110'$ est modulée par les vecteurs $\{\mathbf{d}_1, \mathbf{d}_3, \mathbf{d}_2\}$. Le décodage est effectué par l'algorithme de Viterbi, qui parcourt de proche en proche les chemins dans le trellis pour sélectionner celui le plus probablement reçu. L'algorithme de Viterbi offre, en outre, la possibilité d'éliminer, au fur et à mesure du décodage, les chemins ayant le moins de chance d'être reçus, ce qui permet de réduire significativement le traitement. Dans une modulation en treillis, ce treillis est modifié de sorte que plusieurs arcs peuvent coder un même symbole, chacun de ces arcs étant associé à un des vecteurs du dictionnaire. Le treillis non informé de la figure 3.7 (a) devient celui de la figure 3.7 (b). L'encodage d'un symbole m nécessite de choisir parmi les Q arcs associés à m celui maximisant ses chances de détection. Il choisit donc l'arc correspondant au vecteur du dictionnaire de plus grande corrélation avec le signal reçu. L'encodage de $'110'$ peut devenir en fonction du signal audio la suite de vecteurs $\{\mathbf{d}_1^0, \mathbf{d}_3^1, \mathbf{d}_2^1\}$. L'algorithme de Viterbi est modifié en conséquence pour prendre en compte ces multiples chemins.

3.3.3 Codage informé

3.3.3.1 Insertion par quantification

Les travaux de Chen et Wornell en 2001 marquent le début des recherches sur les fonctions d'insertion par quantification, qualifiées par la suite de ***codage informé***. Ils proposent un nouveau système, la QIM [18] [CW01], dont les performances théoriques sont proches de celle du

18. Quantization Index Modulation

modèle de Costa. Le principe de la QIM est le suivant : la fonction d'insertion correspond à la quantification du vecteur audio $\mathbf{x}$ par un quantificateur dépendant du symbole m à émettre et noté $q_m(\mathbf{x})$:

$$f(\mathbf{x}, m) = q_m(\mathbf{x}). \tag{3.20}$$

Le signal audio tatoué $\mathbf{y}$ devient le représentant, appelé ***atome***, du quantificateur q_m le plus proche de $\mathbf{x}$; il s'agit donc d'une insertion substitutive. Le dictionnaire de canal est cette fois constitué de l'ensemble des atomes associés aux M quantificateurs, définissant chacun un ***réseau***. Sa dimension devient infinie. Le tatouage $\mathbf{t} = \mathbf{y} - \mathbf{x}$ est d'autant moins puissant que la distance entre les atomes voisins est petite. En effet, cette faible distance garantit la présence d'un atome représentant le symbole à émettre (et donc d'un signal audio tatoué) à proximité du signal audio original. Si la distance entre deux atomes voisins est suffisamment petite, la contrainte d'inaudibilité peut être garantie ; elle prend ici la forme d'une mesure de l'erreur quadratique entre le signal audio tatoué et le signal audio original : $E\left[||\mathbf{y} - \mathbf{x}||^2\right]$. La détection consiste cette fois à rechercher l'atome le plus proche du signal reçu. Cet atome est associé à l'un des quantificateurs q_m ; son indice décide le symbole reçu :

$$\hat{m} = \arg \min_{m \in [\![0, M-1]\!]} \left[\hat{\mathbf{y}} - q_m(\hat{\mathbf{y}})\right]^2. \tag{3.21}$$

La performance du détecteur est là encore liée à la distance entre les atomes. En notant d_{min} la plus petite distance séparant deux atomes associés à deux symboles différents, Chen montre que la probabilité d'erreur du système s'exprime sous la forme :

$$P_e = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{d_{min}^2}{8\sigma_p^2}}\right), \tag{3.22}$$

où $\frac{1}{2}\mathrm{erfc}(x) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} \exp\left(-\frac{\theta^2}{2}\right) d\theta$ et σ_p^2 est la variance du bruit de perturbation supposé additif. Cette probabilité devient indépendante du bruit audio, comme préconisé par le schéma de Costa. Mais contrairement à l'inaudibilité qui conduit à choisir des atomes relativement proches, la détection est d'autant meilleure que la distance d_{min} entre les atomes est grande. On retrouve le compromis classique entre inaudibilité et fiabilité de transmission.

La figure 3.8 (b) propose une interprétation géométrique de l'insertion. Deux bits $'0'$ ou $'1'$ peuvent être insérés à l'aide de deux quantificateurs dont les atomes sont représentés par des ronds pour $'0'$ et des carrés pour $'1'$. Tatouer le bit $'1'$ consiste donc à substituer au signal audio l'atome (carré) du quantificateur le plus proche de $\mathbf{x}$. Cet atome devient alors le signal audio tatoué $\mathbf{y}$. Tous les signaux audio situés dans la région d'attraction (représentée par le losange) et devant cacher $'0'$ seront remplacés par cet atome. L'efficacité d'une telle stratégie d'insertion tient à la discontinuité de la région de détection. Le terme discontinu traduit ici l'entrecroisement des régions de détection de chaque message. La région de détection de $'0'$ est la réunion des régions de détection de chaque atome associé à $'0'$. Cette région est discontinue et entrecroisée avec la région de détection du $'1'$, par opposition à la région de détection de '0' pour un schéma de type additif non informé (rappelée figure 3.8 (a)) qui occupe une moitié continue de l'espace de détection. Le choix d'une telle région de détection discontinue garantit de trouver un signal audio tatoué porteur de l'information à émettre et à proximité du signal audio original : inaudibilité (en terme d'erreur quadratique moyenne) et détection correcte sont donc conciliées.

Dans le cas simple d'insertion dans un domaine de faible dimension, la quantification peut être aisément mise en œuvre. Chen propose notamment d'appliquer la QIM à la projection du signal

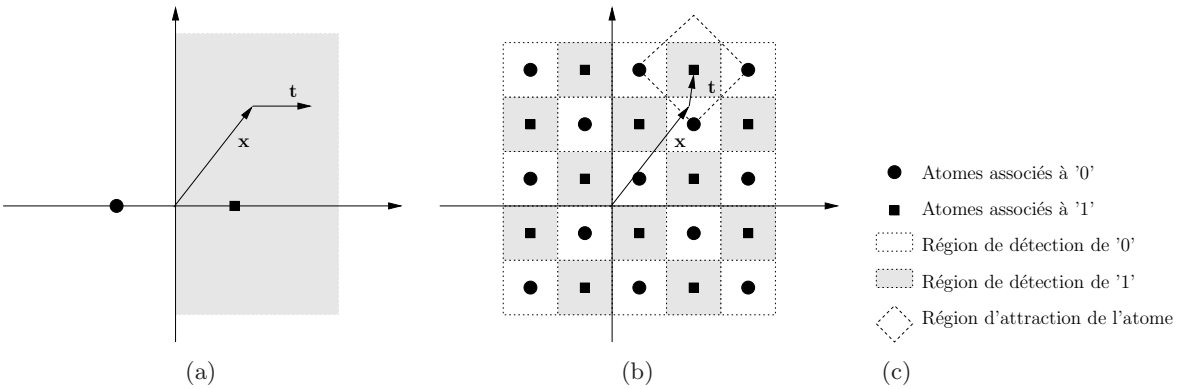


Figure 3.8 – Interprétation géométrique de la QIM [CW01] (b) comparée à celle d'un système additif classique (a) : mise en évidence des régions de détection et du choix du tatouage $\mathbf{t}$ en fonction du signal audio $\mathbf{x}$ lorsque l'information à insérer est $'1'$.

audio sur un vecteur étalé spectralement. Dans ce système appelé STDM [19] [Che00], le vecteur des caractéristiques audio $\mathbf{x}$ à quantifier est donc scalaire. Les quantificateurs le sont également : à l'émetteur, le vecteur audio est quantifié en fonction du symbole à émettre ; au récepteur, il suffit de quantifier le signal audio reçu par l'ensemble des quantifieurs et de comparer la distance entre les points de reconstruction et le signal audio reçu. La mise en œuvre est donc aisée et peu coûteuse en temps de calcul. Dans le cas d'un domaine d'insertion de grande dimension, la mise en œuvre de la quantification et la recherche de l'atome détecté sont nettement plus délicates. Les avancées dans ce domaine exploitent alors les possibilités offertes par les codes correcteurs pour construire un dictionnaire structuré et obtenir un algorithme recherchant l'atome à proximité du vecteur audio efficace.

3.3.3.2 Codes correcteurs d'erreur

Les Codes Correcteurs d'Erreur (CCE) construisent à partir du représentant m à tatouer (codé sous la forme d'un K-uplet binaire) un mot de code de longueur N, qui peut être utilisé comme atome du dictionnaire associé à m. Un algorithme (l'algorithme de Viterbi dans le cas des codes convolutifs par exemple) permet alors, au regard de l'observation d'un code reçu qui est ici le signal audio, de détecter à la fois le mot de code auquel il correspond et le représentant m associé. Différents codes dérivés des CCE sont alors proposés : dans [LGPD03], Le Guelvouit *et al.* proposent l'utilisation de codes poinçonnés basés sur l'algorithme de Viterbi pour le codage et le décodage efficace ; dans [CPGR00], Chou *et al.* proposent un codage de source distribué utilisant les syndromes associés aux CCE.

3.3.3.3 Vers le schéma de Costa

Bien que les fonctions d'insertion par quantification aient des performances très similaires au schéma de Costa, l'expression du tatouage qu'elle propose est légèrement différente de celle du modèle de Costa : la quantification établit $\mathbf{t} = q_m(\mathbf{x})$ tandis que Costa suggère $\mathbf{t} = \mathbf{u} - \alpha\mathbf{x}$.

Différentes variantes des fonctions d'insertion ont donc été construites pour proposer un

19. Spread Transform Dither Modulation

tatouage $\mathbf{t}$ dont le formalisme est celui de Costa. Chen et Wornell proposent notamment la DC-QIM [20] [CW01], pour laquelle la fonction d'insertion devient :

$$f(\mathbf{x}, m) = q_m\left(\mathbf{x}, \frac{\Delta}{\alpha}\right) + (1-\alpha)\left(\mathbf{x} - q_m\left(\mathbf{x}, \frac{\Delta}{\alpha}\right)\right) \tag{3.23}$$

et le tatouage : $\mathbf{t} = \alpha q_m\left(\mathbf{x}, \frac{\Delta}{\alpha}\right) - \alpha\mathbf{x}$, où Δ/α est la quantification du signal audio par un pas $\frac{\Delta}{\alpha}$. Cette fonction, en plus de sa ressemblance avec le modèle de Costa, permet d'augmenter la distance minimale entre deux atomes d'un facteur $1/\alpha^2$ et donc d'augmenter la robustesse du système aux bruits additifs, tout en compensant l'augmentation de la distorsion perceptuelle par l'ajout d'une fraction $(1-\alpha)$ de l'erreur de quantification.

Eggers *et al.* [EBTG03] proposent quant à eux une application du schéma de Costa au cas scalaire. Son schéma de tatouage, le SCS [21], utilise un unique quantificateur scalaire de pas Δ. Le symbole à insérer est alors pris en compte par l'intermédiaire d'un offset décalant les atomes de quantification. La fonction d'insertion devient :

$$f(\mathbf{x}, m) = \mathbf{x} + \alpha\mathbf{t}, \text{ avec } \mathbf{t} = q\left(\mathbf{x} - \Delta\frac{m}{M}\right) - \left(\mathbf{x} - \Delta\frac{m}{M}\right), \tag{3.24}$$

où M est le nombre de symboles possibles. Toutefois, bien que le formalisme de Eggers *et al.* soit identique à celui de Costa, les auteurs montrent expérimentalement que la force d'incrustation α adéquate diffère de Costa. Elle prend la forme :

$$\alpha = \sqrt{\frac{\sigma_t^2}{\sigma_t^2 + 2,71\sigma_p^2}}, \tag{3.25}$$

où σ_t^2 est la variance du tatouage et σ_p^2 celle du bruit, alors que l'expression établie par Costa pour atteindre la capacité d'insertion maximale est $\alpha = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_p^2}$.

3.3.3.4 Robustesse aux distorsions

Les techniques de codage informé, parce qu'elle mettent en œuvre une insertion proche du modèle de Costa, sont très performantes lorsque le canal ne subit pas ou peu de perturbations : la capacité d'insertion est très élevée et l'on peut concevoir un système de tatouage dont les TEBs même à débit élevé (de l'ordre du kbit/s) sont très faibles.

Pourtant, ces techniques n'atteignent pas encore les contraintes de robustesse exigées pour le système de tatouage. Dès lors qu'un simple facteur d'amplitude est appliqué au signal audio tatoué, les TEBs augmentent considérablement ; ces systèmes sont peu robustes aux perturbations de canal. Deux approches ont donc été envisagées pour palier à ce manque de robustesse. Les premières consistent à choisir un domaine d'insertion invariant aux perturbations ; mais l'on connaît la difficulté de trouver un tel domaine. Les secondes introduisent un module d'estimation de l'attaque : l'objectif est d'estimer, au récepteur, la perturbation effectuée par le canal (le facteur d'amplitude [EBG01] par exemple), pour inverser la distorsion avant de procéder à l'extraction. Bien que des avancées aient été faites dans le domaine, on ne dispose pas à notre

20. Distorsion Compensation Quantization Index Modulation
21. Scalar Costa Scheme

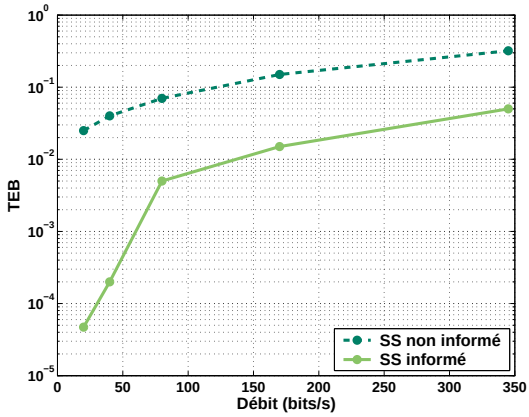


Figure 3.9 – Performances du système de tatouage informé additif par étalement de spectre dans le domaine temporel de Cvejic et Seppänen (extrait de [CS03]).

connaissance d'une implémentation de systèmes de codage informé qui répondent à toutes les attentes en matière de robustesse, et notamment aux perturbations désynchronisantes. Des travaux théoriques sur l'influence de la désynchronisation [ZBDre] (modélisée par l'ajout de retards aléatoires dans le signal audio) confirment l'efficacité des stratégies substitutives en terme de débit de transmission sur les stratégies additives ; mais la question de leur mise en œuvre effective reste encore d'actualité.

3.3.4 Performances

Évaluer les performances des systèmes avec insertion informée révèle des difficultés similaires aux cas des systèmes non informés. Le respect de la contrainte d'inaudibilité est une limite majeure : en effet, dans le cas des fonctions d'insertion additives informées, cette contrainte est volontairement négligée au profit de la robustesse, conformément à la stratégie de robustesse constante proposée par Miller : l'inaudibilité n'est mesurée qu'après insertion du tatouage. Dans le cas du codage informé, cette contrainte est prise en compte par l'erreur quadratique moyenne entre le signal audio tatoué et le signal original. Or, ce critère d'EQM est mal adapté au contexte du tatouage. La littérature ne fait donc pas état de courbes de TEB en fonction du débit à "inaudibilité constante". Pour donner un ordre de grandeur des performances de la littérature nous ne citerons donc que l'article de Cvejic et Seppänen [CS03] qui propose une implémentation de l'ISS de Malvar dans laquelle le tatouage est mis en forme spectralement par un modèle psychoacoustique pour garantir son inaudibilité. Les performances établies par les auteurs sont données figure 3.9. Elles mettent en évidence le gain en terme de TEB des systèmes informés sur les systèmes non informés, mais ne permettent pas d'évaluer l'impact de la stratégie sur l'inaudibilité du système.

3.4 Spécificités des systèmes pour la synchronisation

Parmi l'ensemble des perturbations auxquelles un système de tatouage se doit d'être robuste, les opérations désynchronisantes sont sans doute les plus dégradantes pour la détection du

tatouage. Comme nous l'avons évoqué précédemment, cette désynchronisation apparaît pour deux types d'applications : le transfert d'information via l'onde acoustique et la radiodiffusion. Les altérations produites sur le signal audio peuvent introduire :

- un simple retard entre le signal audio tatoué à l'émission et sa version enregistrée à la réception, dans le cas d'opération de filtrage par exemple,
- un réechantillonnage du signal audio tatoué suite à des conversions numérique-analogique et analogique-numérique dans le cas d'un transfert du signal audio tatoué entre 2 PCs via l'onde acoustique,
- une opération de dilatation ou de compression du signal audio pour des applications de type radio diffusion.

Si le domaine d'insertion n'est pas invariant par l'opération désynchronisante, ces perturbations entraînent la perte de la localisation des symboles voire la perte de la durée N_s des symboles : ce phénomène est particulièrement problématique dans les systèmes de tatouage par étalement de spectre réalisant l'addition du tatouage au signal audio dans le domaine temporel. En effet, le décalage d'un seul échantillon sur la localisation du symbole émis dans le signal reçu suffit à dégrader significativement les performances du démodulateur par corrélation. Un paramètre de dérive peut alors être défini pour caractériser la différence entre le temps symbole N_s à l'émission et à la réception.

Les techniques à mettre en œuvre pour permettre la resynchronisation de ces systèmes sont très différentes des stratégies d'insertion informées, établies précédemment pour garantir la robustesse du système à l'ajout d'un bruit additif. Elles s'attachent à la mise en place d'un système de **synchronisation de rythme**, qui peut être dissocié de la stratégie d'insertion.

Dans cette section, nous présentons les différents mécanismes de synchronisation, proposés par la théorie des communications analogiques et numériques et leurs applications à la conception de systèmes de tatouage audio. Ces dernières ont été majoritairement développées pour des chaînes de communication additives dans le domaine temporel.

3.4.1 Modèle du schéma de tatouage

Nous considérons dans cette section un modèle simplifié de la chaîne de tatouage, présenté figure 3.10. L'information à émettre b est supposée binaire, le tatouage est construit à l'aide d'un unique vecteur de dictionnaire $\mathbf{d}$ de sorte que $\mathbf{t} = (2b-1)\mathbf{d}$. L'insertion est effectuée après découpage du signal audio $x(n)$ en fenêtre d'analyse de durée N_s appelée temps symbole. Sur la l-ième fenêtre, le signal audio tatoué est construit en ajoutant le tatouage porteur de b au signal audio original :

$$y(n) = x(n) + (2b-1)\mathbf{d}[n - lN_s], \tag{3.26}$$

où $\mathbf{d}[i]$ désigne la i-ième composante de $\mathbf{d}$. La détection est effectuée en calculant la corrélation entre le signal reçu $\hat{y}(n)$ découpé en fenêtre d'analyse de taille N_s et le vecteur du dictionnaire. Le signe de la corrélation indique le bit reçu. Dans la suite du développement, nous nous placerons lors de la détection et de la synchronisation du l-ième symbole d'information, i.e. sur le l-ième temps symbole. Elle concerne théoriquement la fenêtre du signal reçu allant de l'échantillon lN_s à l'échantillon $(l+1)N_s$, que nous noterons $\hat{\mathbf{y}}_{N_s}(lN_s)$: fenêtre du signal $\hat{y}(n)$ débutant à l'échantillon lN_s et de durée N_s.

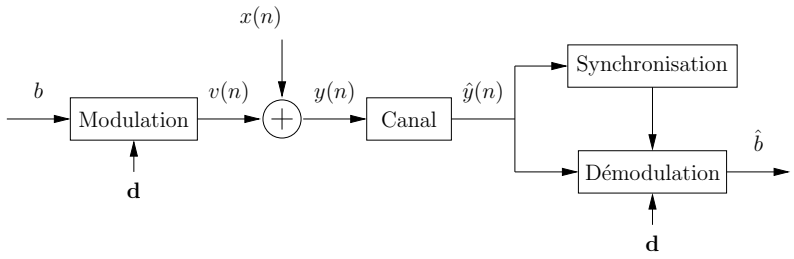


Figure 3.10 – Schéma simplifié du système de tatouage pour la mise en évidence des mécanismes de synchronisation.

3.4.2 Synchronisation des chaînes de communication analogique

En matière de synchronisation de rythme dans les chaînes de communication analogique, les systèmes les plus couramment utilisés sont :

- les synchroniseurs de type ***early-late gate*** [Pro01] : Pour le l-ième temps symbole, trois corrélations sont effectuées, de $\hat{\mathbf{y}}_{N_s}(lN_s + p)$ pour $p \in \{-1, 0, 1\}$ avec la forme d'onde du dictionnaire $\mathbf{d}$. L'échantillon p_{opt} maximisant la corrélation :

$$p_{opt} = \arg\max_{p} \hat{\mathbf{y}}_{N_s}(lN_s + p)^t s, \tag{3.27}$$

  indique l'instant de début du bit en cours de détection (en avance par rapport à la localisation théorique si la corrélation est maximale pour $\hat{\mathbf{y}}_{N_s}[lN_s - 1]$ et en retard si la corrélation est maximale pour $\hat{\mathbf{y}}_{N_s}[lN_s + 1]$). Le signe de la corrélation donne la valeur du bit reçu. Ce type de synchronisation est efficace à la condition que la durée du temps symbole à la réception soit peu modifié par rapport à l'émission.

- la ***méthode de la raie spectrale*** [Bro97] : La construction du signal, par répétition des vecteurs du dictionnaire tous les temps symbole, introduit un pic à la fréquence $\frac{F_e}{N_s}$ dans le spectre du signal. Cette méthode consiste donc à estimer le temps symbole N_s à la réception en détectant le pic fréquentiel par une opération de filtrage autour de la fréquence $\frac{F_e}{N_s}$.

Ces mécanismes de synchronisation sont efficaces lorsque la dérive entre l'émetteur et le récepteur est relativement faible ; dès que la dérive devient plus élevée (i.e. que le temps symbole au récepteur est très différent de celui à l'émetteur), le mécanisme de synchronisation risque de décrocher.

3.4.3 Synchronisation des chaînes de communication numérique

En matière de synchronisation pour les chaînes de communication numérique, une importante partie de la littérature est dédiée aux mécanismes de synchronisation pour les systèmes par étalement de spectre [Pro01].

L'implémentation des systèmes par étalement de spectre utilise généralement des horloges très précises. La fréquence de réception du signal est donc peu différente de la fréquence d'émission. Le mécanisme de synchronisation est généralement scindé en 2 étapes : une phase d'acquisition

initiale et une phase de poursuite. L'acquisition initiale a pour but d'estimer le retard à l'origine entre le signal reçu et le signal émis. Elle utilise une séquence d'initialisation, connue du récepteur. Le récepteur exploite une technique de corrélation par fenêtre glissante pour détecter la séquence d'initialisation dans le signal reçu. Dès que la corrélation devient supérieure à une valeur seuil, préalablement fixée, la séquence d'initialisation est détectée. De fait, un bruit de canal peut conduire à des fausses alarmes de détection ; un circuit additionnel peut alors être utilisé en recours pour détecter ces fausses alarmes. Une fois la séquence d'initialisation détectée, le récepteur passe en mode de détection des données. Il s'agit de la seconde étape, celle de poursuite ou de synchronisation fine. Le but est ici de synchroniser le récepteur sur chaque bit d'information. Deux techniques sont couramment utilisées : la Delay-Locked Loop (DLL) et la Tau-Dither Loop (TDL). Toutes deux sont des adaptations du early-late gate synchronizer, la TDL étant plus rapide et moins coûteuse en temps de calcul que la DLL. Elles utilisent donc toutes les deux une méthode de corrélation par fenêtre glissante pour détecter la synchronisation.

Les systèmes de synchronisation proposés dans la littérature se basent donc principalement sur les propriétés d'autocorrélation des formes d'onde utilisées pour la transmission. Ils font appel à des séquences pseudo-aléatoires spécifiques, parmi lesquelles on dénombre notamment les séquences Gold [McE87], qui garantissent une fonction d'autocorrélation proche d'un dirac, ou les séquences CEC [22] proposées pour l'UMTS [23] [RJ01]. Ces dernières, en plus des propriétés de leurs fonctions d'autocorrélation, offrent l'avantage que la somme de fonctions d'autocorrélation de deux séquences est un dirac. Ceci permet la mise en œuvre de mécanismes de synchronisation rapides et peu coûteux en temps de calcul.

3.4.4 Synchronisation dans les systèmes de tatouage audio

Quelques mécanismes de synchronisation pour les systèmes de tatouage audio ont été présentés dans la littérature. Ils peuvent être classés en 2 catégories, celles des mécanismes exploitant les propriétés de resynchronisation d'un tatouage étalé spectralement et celles basées sur des points d'intérêt.

3.4.4.1 Synchronisation grâce à l'étalement de spectre

Cette première catégorie de mécanisme de synchronisation regroupe les systèmes exploitant les propriétés de corrélation des séquences pseudo-aléatoires.

Le principe initial est le suivant : des patterns de synchronisation sont insérées tout au long du signal de tatouage porteur d'une information binaire [Gom02, Bar02]. Ces patterns résultent du tatouage d'une forme d'onde spécifique étalée spectralement de durée finie N_p et connue du récepteur. La détection consiste alors à localiser ces patterns de synchronisation à l'aide d'une technique de corrélation par fenêtre glissante. En effet, la pattern étant étalée spectralement, la corrélation n'est maximale que lorsque la fenêtre d'analyse du signal reçu (sur laquelle est calculée cette corrélation) coïncide avec l'insertion de la pattern de synchronisation. L'inconvénient d'une telle méthode tient au fait que l'insertion des patterns diminue fortement le débit de transmission utile. Pour palier à cet inconvénient, Gomes [Gom02] propose une variante de ce mécanisme dans laquelle l'insertion des patterns de synchronisation est effectuée parallèlement à la transmission de l'information binaire. Le signal de tatouage est la somme de deux composantes : la première

22. Concatenated Extended Complementary
23. Universal Mobile Telecommunication System

résulte de l'insertion de l'information et la seconde d'un tatouage de synchronisation (insertion des patterns de synchronisation). La détection du tatouage de synchronisation localise le début de chaque symbole d'information, qui peut ensuite être détecté. La synchronisation est finalement réalisée grâce à l'étalement de spectre de la pattern. Pour des systèmes de tatouage où l'insertion de l'information s'effectue à l'aide de vecteurs étalés spectralement et la détection est basée sur un détecteur par corrélation, il est donc possible de s'affranchir de l'introduction des patterns. La synchronisation et la détection sont effectuées conjointement en recherchant par une technique de corrélation avec fenêtre glissante le maximum de corrélation. Cette idée sera utilisée dans le système de Kirovski [KM03] décrit par la suite.

Ces premières méthodes sont relativement efficaces à condition que l'opération désynchronisante soit un retard ou une modification de l'échantillonnage de faible dérive (inférieure à une dizaine d'échantillons par seconde). En effet, pour des opérations désynchronisantes plus complexes, telles des dilations ou des compressions de la durée du signal, la durée du temps symbole N_s tout comme celle de la pattern de synchronisation N_p dans le signal reçu est modifiée, ce qui perturbe fortement la technique de corrélation par fenêtre glissante.

Deux approches peuvent alors être envisagées. La première est naturellement de choisir un domaine d'insertion invariant par dilatation temporelle de sorte que l'opération de désynchronisation ne modifie que la localisation des symboles mais pas leur durée. Le domaine envisagé est souvent le domaine fréquentiel (MDCT [24], DCT [25], DWT [26]) [WHHS05, WHHS04]. La seconde consiste à calculer des versions dilatées des formes d'onde (pattern et vecteurs du dictionnaire d'émission) selon différentes valeurs de dilatation et à mettre en œuvre la technique de corrélation par fenêtre glissante pour ces versions dilatées [BvdVL04].

Dans ce domaine, Kirovski et Malvar [KM03] proposent sans doute l'un des systèmes les plus aboutis : leur système permet de détecter des modifications de l'échelle aussi bien temporelle que fréquentielle en utilisant une forme d'onde à deux dimensions (l'une temporelle et l'autre fréquentielle). Cette forme d'onde joue à la fois le rôle de vecteur support de l'information $\mathbf{d}$ et de pattern de synchronisation. Elle est insérée dans le domaine fréquentiel. La technique de corrélation par fenêtre glissante est utilisée pour la synchronisation : des valeurs de la corrélation entre le signal reçu (pour différents fenêtrages dans le temps) et des versions dilatées de la forme d'onde (pour différents facteurs de dilatation) sont calculées. La corrélation maximale permet alors d'estimer le facteur de dilatation temporelle en localisant dans le temps la fenêtre où est insérée la forme d'onde, le facteur de dilatation fréquentielle en déterminant la version dilatée la plus ressemblante à la forme d'onde reçue et le bit reçu. Les tests que nous avons réalisés sur ce système tendent à montrer que le mécanisme de synchronisation est efficace ; il est malheureusement très coûteux en temps de calcul, car la recherche des facteurs de dilation est effectuée de manière exhaustive (i.e. en analysant tous les décalages possibles dans le temps et toutes les dilatations en fréquences possibles) sur chaque temps symbole. Ce coût empêche pour l'instant l'application de cette méthode à des systèmes temps-réel.

3.4.4.2 Synchronisation par points d'intérêt

Les systèmes de synchronisation de la seconde catégorie extraient du signal audio différents points d'intérêt, qui définissent une cartographie du signal audio tatoué. Chaque point d'intérêt

24. Modulated Discret Cosinus Transform
25. Discrete Cosinus Transform
26. Discrete Wavelet Transform

marque la région du signal audio où est insérée l'information. L'extraction de cette "carte" à la réception permet d'estimer puis d'inverser l'opération désynchronisante et de localiser la région du signal audio où a été inséré le tatouage.

Ces points d'intérêt peuvent être relatifs à l'énergie du signal [HS02], à l'amplitude en sortie d'un filtrage non linéaire [LMBL04] inspiré du tatouage d'images [Bas00] ou à la détection du tempo [KA03]. Les défauts majeurs de ces systèmes sont sans doute leurs débits variables : les points d'intérêts n'étant pas disposés à intervalles réguliers, la quantité d'information transmise par seconde est dépendante du signal audio. Ce débit variable peut ne pas être compatible aux applications cibles auxquelles se destine le tatouage.

Conclusion

Le tatouage des signaux audio relève de deux grandes classes applicatives : les applications à vocation sécuritaire (protection de la propriété intellectuelle par exemple) et les applications de transmission de données (pour l'augmentation du contenu audio par exemple), auxquels s'attache ce manuscrit.

Dans ce contexte particulier, le système de tatouage doit satisfaire à quatre contraintes définissant la performance du système : l'inaudibilité de la transmission, un débit de transmission le plus élevé possible pour une fiabilité de transmission la meilleure possible, la robustesse à différentes perturbations licites, classiquement appliquées au signal audio, et le coût en temps de calcul permettant l'application temps-réel.

Ces critères de performances peuvent être mesurés par un protocole expérimental précis que nous avons établi. Ce protocole exploite un algorithme objectif de mesure de la différence perceptible, l'algorithme PEAQ [Uni01], entre le signal audio original et sa version tatouée pour jauger de la transparence du tatouage ; il propose pour évaluer la fiabilité de transmission de mesurer du TEB moyen obtenu par tatouage d'un corpus de 20 signaux audio de style divers. Ce TEB sert également à évaluer la robustesse du système aux perturbations. Ces perturbations sont choisies pour être représentatives des opérations licites pouvant être effectuées sur un signal audio et seront réalisées en utilisant l'outil Stirmark Audio [Lan]. Il comprend des conversions de format (compression MPEG), des modifications d'échelle (d'amplitude, mais également des opérations de dilation ou de contraction de l'échelle des temps), des filtrages, etc... Ce protocole propose également un critère de mesure du coût en temps de calcul des systèmes de tatouage. Il sera repris tout au long de ce manuscrit pour évaluer les performances des systèmes proposés.

De nombreuses solutions sont proposées dans la littérature pour résoudre le problème du tatouage. Une partie se base sur l'analogie entre chaîne de tatouage et chaîne de communication pour construire des systèmes de tatouage relativement efficaces mais dont les performances sont inévitablement limitées par le signal audio. Pour palier à ce problème, des stratégies d'insertion informées ont été introduites. Une insertion est dite informée si la technique de construction du tatouage exploite la connaissance *a priori* du signal audio à l'émetteur. Un tatouage inaudible et de robustesse maximale aux perturbations peut donc théoriquement être construit. L'efficacité théorique de ces stratégies informées sur les performances du système de tatouage a d'ores et déjà été établie. Elle nous incite de fait à construire un système de tatouage exploitant une technique d'insertion informée. Deux grandes familles de stratégie d'insertion informée se distinguent dans la littérature : les techniques de codage informé (dont fait notamment parmi la QIM de Chen [Che00] ou le SCS de Eggers [EBTG03]) et les stratégies d'insertion additive informée (telle que l'ISS de Malvar et Florencio [MF03] par exemple). Les techniques de codage informé bien que permettant d'atteindre des débits de transmission élevés se révèlent peu robuste aux perturbations (notamment les perturbations désynchronisantes). Les fonctions d'insertion additives

informées, plus robustes aux perturbations, sont majoritairement construites pour satisfaire la contrainte de robustesse souvent au détriment (ou avec une mauvaise prise en compte) de la contrainte d'inaudibilité. L'objet de cette thèse est donc de proposer un système de tatouage de type additif informé qui garantisse en premier lieu le respect de la contrainte d'inaudibilité et choisisse ensuite un tatouage de robustesse maximale aux perturbations.

Dans cet objectif, nous nous proposons dans un premier temps d'établir un système de tatouage non informé, dont la construction fait l'objet des chapitres à venir.

Deuxième partie

Un système de tatouage audio

Introduction

L'objectif de cette partie est de présenter la conception d'un système de tatouage audio, en exploitant l'analogie entre les systèmes de tatouage et les chaînes de communication numériques. Ce principe n'est pas nouveau et le système présenté dans cette partie s'appuie fortement sur les travaux de L. de C. T. Gomes [Gom02], réalisés au cours de son doctorat à l'UFR de Mathématiques et Informatique de l'Université Paris V et à l'ENST. Une importante partie de ce travail s'est donc attachée à formaliser puis à justifier les étapes de la conception du système. Ces justifications s'appuient aussi bien sur les principes de la théorie des communications numériques [Pro01, Cha03], sur des éléments de psychoacoustique [ZF81] que sur des mesures expérimentales permettant la paramétrisation du système de tatouage.

Le système proposé est conçu sous la forme d'une chaîne de communication, reprenant les trois étapes classiques que sont :

- l'émetteur, dont le rôle est de construire le signal audio tatoué $y(n)$ en modulant l'information à émettre. Cette modulation conduit au signal de tatouage $t(n)$ inaudible, inséré par addition au signal audio hôte $x(n)$.
- le canal, siège des perturbations que peut subir le signal audio tatoué et que l'on supposera dans toute cette partie non désynchronisantes,
- le récepteur qui extrait du signal audio reçu l'information cachée.

Le premier chapitre de cette partie se propose de rappeler les outils de la théorie des communications numériques nécessaires à la conception et à l'évaluation des performances théoriques de la chaîne de tatouage.

Le second chapitre de cette étude s'attache à justifier la conception du système de tatouage sous la seule hypothèse que le signal audio tatoué $y(n)$ est construit en ajoutant le signal de tatouage $t(n)$ au signal audio $x(n)$ dans le domaine temporel. Il met à profit les éléments de la psychoacoustique pour déterminer les conditions à satisfaire par le tatouage pour être inaudible en présence du signal audio. Ces conditions prennent la forme d'un seuil de masquage, paramétré par un facteur de gain α et s'expriment concrètement par une mise en forme spectrale du signal de tatouage. Les outils de la théorie des communications numériques rappelés précédemment sont ensuite exploités pour construire un récepteur sous-optimal adapté au bruit prédominant de la chaîne de tatouage, à savoir le signal audio, et aux spécificités de l'émetteur.

Le système de tatouage est défini à un certain nombre de paramètres près, qu'il convient de fixer : ces paramètres sont liés aux filtres de la chaîne de tatouage, à la bande passante de la chaîne mais également au choix du facteur de gain contrôlant l'inaudibilité. Pour une transparence du tatouage donnée, le choix du facteur de gain adapté s'avère très dépendant du signal audio. Nous proposons donc l'introduction d'un module de contrôle adaptatif de l'inaudibilité, au principe novateur, qui permet de choisir α spécifiquement en fonction du signal audio pour atteindre une

inaudibilité donnée. L'ensemble de ces paramétrisations fait l'objet du troisième chapitre.

Le dernier chapitre est consacré à l'étude des performances du système de tatouage au travers des quatre critères proposés par le protocole expérimental : l'inaudibilité, le TEB en fonction du débit de transmission pour un canal sans perturbation, la robustesse aux perturbations et le temps de calcul. Les perturbations seront supposées non désynchronisantes, l'étude du mécanisme de synchronisation étant reporté au chapitre 11. Les performances de la chaîne sur des signaux réels sont présentées, discutées et comparées à un système de l'état de l'Art.

Par la suite, le système construit dans cette partie sera appelé ***système non informé de référence avec module de contrôle de l'inaudibilité***.

Chapitre 4

Éléments de théorie des communications numériques

La conception du système de tatouage se base principalement sur l'hypothèse que le signal de tatouage $t(n)$ est ajouté au signal audio hôte $x(n)$ directement dans le domaine temporel. Le signal audio tatoué $y(n)$ s'écrit donc simplement :

$$y(n) = x(n) + t(n). \tag{4.1}$$

Ce formalisme établit immédiatement le lien entre système de tatouage et chaîne de communications numériques. Le signal de tatouage $t(n)$ est le signal utile, porteur de l'information à transmettre, tandis que le signal audio $x(n)$ est le bruit caractérisant le canal de transmission.

La théorie des communications numériques fournit les pré-requis à la construction d'une chaîne de tatouage, qui font de fait l'objet de ce chapitre. En supposant l'information à tatouer être une séquence binaire, elle formalise la modulation dont est déduit le signal $t(n)$ et les conditions de transmission (rapport signal à bruit, débit de transmission, etc...). Elle définit également le récepteur optimal lorsque le bruit audio $x(n)$ peut être supposé blanc Gaussien. Pour finir, elle établit les performances théoriques de la chaîne de communication en terme de probabilité d'erreur de transmission. Ces trois points, modulation, réception et performances théoriques, sont détaillées dans les sections à venir.

Pour rappeler ces éléments de théorie des communications numériques [Pro01], nous nous référerons tout au long de ce chapitre à la chaîne de communication numérique standard, dont le schéma de principe est présenté figure 4.1. Cette chaîne est divisée en trois étapes classiques, que sont l'émetteur construisant le signal $t(n)$, le canal supposé ici réduit au bruit audio $x(n)$ et le récepteur.

4.1 Message numérique

L'information à tatouer, ou ***message numérique***, est supposée être une suite i.i.d. de L K-uplet binaires :

$$\left(\left(b_0[0], b_0[1], ..., b_0[K-1]\right), ..., \left(b_{L-1}[0], b_{L-1}[1], ..., b_{L-1}[K-1]\right)\right) \tag{4.2}$$

à valeurs dans $\left\{\{0,1\}^K\right\}^L$. La procédure d'encodage, préliminaire à l'étape de modulation, permet d'interpréter chaque K-uplet binaire $(b_l[0], b_l[1], ..., b_l[K-1])$ de cette suite de deux

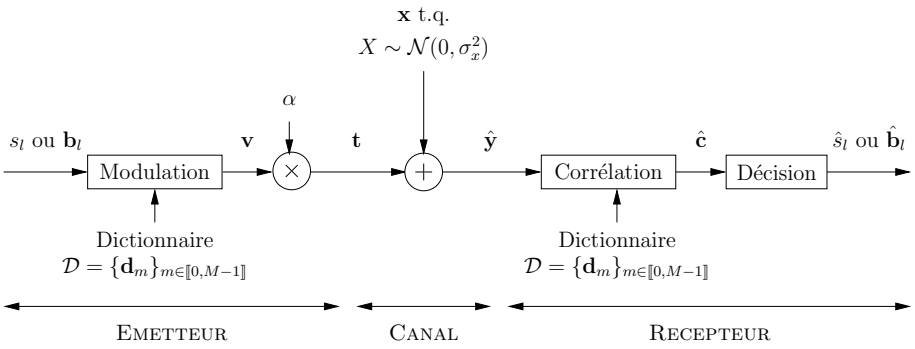


Figure 4.1 – Modèle synthétique de la chaîne de tatouage.

façons :

- soit sous la forme d'un ***symbole*** s_l (entier) à valeur dans l'alphabet $\{0, \ldots, 2^K - 1\}$:

$$s_l = \sum_{k=0}^{K-1} b_l[k] 2^k, \tag{4.3}$$

- soit sous la forme d'un ***vecteur*** $\mathbf{b}_l$:

$$\mathbf{b}_l = [b_l[0], b_l[1], \ldots, b_l[K-1]]^t. \tag{4.4}$$

4.2 Modulation

L'étape de modulation consiste à réaliser une association bijective entre un message numérique parmi les 2^{KL} possibles et un signal numérique, le ***signal modulé*** noté $v(n)$. Cette étape de modulation est classiquement caractérisée par la donnée de M signaux à support temporel N_s fini, appelés ***formes d'onde*** et notés $\{d_m(n)\}_{m \in [\![0, M-1]\!]}$. Pour simplifier le formalisme de la modulation, nous définissons la matrice de dimension $N_s \times M$, appelée ***dictionnaire d'émission***, par :

$$\mathcal{D} = \{\mathbf{d}_m\}_{m \in [\![0, M-1]\!]}, \text{ avec } \forall m \in [\![0, M-1]\!], \mathbf{d}_m = [d_m(0), d_m(1), ...d_m(N_s - 1)]^t. \tag{4.5}$$

4.2.1 Signal modulé

Le signal modulé $v(n)$ est construit à partir de ce dictionnaire d'émission en fonction de la forme prise après encodage par le message numérique. $v(n)$ s'exprime :

- pour un message numérique sous la forme d'une suite de symboles,

$$v(n) = \sum_{l=0}^{L-1} d_{s_l}(n - lN_s), \tag{4.6}$$

ce qui impose au dictionnaire d'émission d'être de dimension $M = 2^K$. Ainsi, sur chaque intervalle de temps $[lN_s; (l+1)N_s - 1]$, $v(n)$ coïncide avec le vecteur du dictionnaire dont l'indice est le symbole s_l inséré. Cette modulation est qualifiée de ***modulation*** M***-aire***.

- pour un message numérique sous forme d'une suite de vecteurs,

$$v(n) = \sum_{l=0}^{L-1} \sum_{k=0}^{K-1} (2b_l[k] - 1)\, d_k(n - lN_s), \tag{4.7}$$

  à condition que le dictionnaire d'émission soit de dimension $M = K$. Cette expression reflète à la fois une ***modulation CDMA*** [Vit95], utilisée en communication numérique pour véhiculer plusieurs informations dans un même canal, ou la somme de modulations BPSK utilisant K porteuses (éventuellement disjointes). Dans ces rappels de théorie des communications numériques, nous nous limiterons au cas des modulations CDMA.

Dans cette chaîne, le signal de tatouage $t(n)$ est obtenu en pondérant le signal modulé $v(n)$ par le facteur d'amplitude α. Ce facteur d'amplitude règle l'énergie consacrée à la transmission de chaque symbole ou K-uplet binaire et le RSB, rapport entre la puissance du signal utile $t(n)$ et du bruit de canal $x(n)$.

4.2.2 Débit de transmission

La durée des vecteurs du dictionnaire, N_s, est le ***temps symbole***, c'est à dire le temps en échantillon consacré à la transmission de chaque K-uplet binaire. Ce paramètre influe sur le ***débit de transmission*** R, défini comme le nombre de bits d'information transmis par unité de temps. La chaîne étant numérique, cette unité de temps est liée à la fréquence d'échantillonnage F_e. Le débit de transmission en bits par seconde s'exprime finalement par :

$$R = \frac{F_e}{N_b} = \frac{KF_e}{N_s}, \tag{4.8}$$

où $N_b = \dfrac{N_s}{K}$ est équivalent au nombre d'échantillons consacrés à la transmission d'un bit.

4.3 Récepteur optimal

Le récepteur optimal est le procédé qui permet de retrouver parmi tous les messages numériques possibles celui qui minimise le nombre d'erreurs au regard de l'observation du signal reçu $y(n)$ sur une durée $N = LN_s$. Ce signal, étant donné le modèle de la chaîne considéré, s'écrit sous la forme :

$$y(n) = x(n) + t(n) = x(n) + \alpha v(n). \tag{4.9}$$

Ce récepteur est parfaitement défini par la théorie de communication numérique [Pro01] dans le cas d'un canal BBAG. Dans cette section, nous supposerons donc le bruit audio $x(n)$ être la réalisation d'un processus aléatoire $X(n)$ blanc Gaussien de loi $\mathcal{N}(0, \sigma_x^2)$, où σ_x^2 est la variance de $X(n)$.

4.3.1 Principe de la décision statistique

Notons $\mathbf{v}$, $\mathbf{x}$ et $\mathbf{y}$ les vecteurs d'observations de $v(n)$, $x(n)$ et $y(n)$ sur un horizon de $N = LN_s$ échantillons. Le signal reçu $\mathbf{y} = \mathbf{x} + \alpha\mathbf{v}$ est la somme d'une composante déterministe $\alpha v(n)$ et

d'une composante aléatoire $X(n)$ dont est issue la réalisation $x(n)$. Étant construit à partir d'un dictionnaire d'émission contenant M vecteurs, $v(n)$, observé sur une durée N, peut prendre 2^{KL} états, correspondants aux 2^{KL} messages numériques possibles. Ces états seront modélisés par une variable aléatoire $V \in \{0, \ldots, 2^{KL} - 1\}$. Le principe du récepteur optimal est donc de déterminer l'état du signal $v(n)$ (et le message numérique qui lui est associé) qui minimise le nombre d'erreurs de transmission étant donnée l'observation $y(n)$.

Supposons connue la loi de probabilité du signal observé $\mathbf{y}$ modélisé par un vecteur aléatoire $\mathbf{Y}$ de longueur N conditionnellement aux états possibles du signal $\mathbf{v}$ sur lesquels portent la décision. Ce vecteur est à valeur dans un espace $\mathfrak{Y}$. De la loi de probabilité peut être déduite notamment la densité de probabilité $p_{\mathbf{Y}|V=m}(\mathbf{y})$ de $\mathbf{Y}$ sachant l'état $V = m$ du signal modulé. La décision statistique consiste à définir une partition $\{\mathfrak{Y}_0, ..., \mathfrak{Y}_{2^{KL}}\}$ de l'espace $\mathfrak{Y}$ de telle sorte que si l'observation $\mathbf{y}$ appartient à $\mathfrak{Y}_m$, l'état du signal modulé $\mathbf{v}$ est décidé être m. Cette partition est choisie pour minimiser le nombre d'erreurs de transmission, c'est à dire pour maximiser la probabilité d'une décision correcte quant à l'état de $\mathbf{v}$. Cette probabilité, notée P_c, peut s'exprimée par :

$$P_c = \sum_{m=0}^{2^{KL}} \text{Prob}(V = m) \int_{\mathfrak{Y}_m} p_{\mathbf{Y}|V=m}(\mathbf{y}) d\mathbf{y}, \tag{4.10}$$

où $\text{Prob}(V = m)$ est la probabilité que V soit égal à m.

Le message numérique étant supposé être une suite i.i.d de K-uplets binaires, la loi de probabilité de V est uniforme. Dans ce cas particulier, la maximisation de la probabilité d'une décision correcte conduit à la règle du ***maximum de vraisemblance*** qui permet de définir la partition de l'espace d'observation. Cette règle énonce que l'état $\hat{m}$ décidé du signal modulé $v(n)$ est celui qui maximise les densités de probabilité $p_{\mathbf{Y}|V=m}(\mathbf{y})$:

$$\hat{m} = \arg\max_m p_{\mathbf{Y}|V=m}(\mathbf{y}). \tag{4.11}$$

En effet, maximiser $\int_{\mathfrak{Y}_m} p_{\mathbf{Y}|V=m}(\mathbf{y}) d\mathbf{y}$ sur les états m possibles sachant $p_{\mathbf{Y}|V=m}(\mathbf{y}) > 0$ revient à maximiser $p_{\mathbf{Y}|V=m}(\mathbf{y})$.

L'expression du détecteur optimal nécessite donc d'expliciter la loi de probabilité du signal reçu. Cette loi dépend elle même de la nature du bruit audio, que nous avons dans cette section supposé BBAG de loi $\mathcal{N}(0, \sigma_x^2)$.

4.3.2 Métriques de décision

Ayant posé $x(n)$ blanc, Gaussien, centré et de variance σ_x^2, le vecteur reçu $\mathbf{y}$, observation de $y(n)$ sur N échantillons, peut alors être modélisé par un vecteur aléatoire $\mathbf{Y}$ dont les N composantes sont N variables aléatoires indépendantes. La densité de probabilité de $\mathbf{Y}$ conditionnellement à l'état du signal modulé $\mathbf{v}$ est donnée par :

$$p_{\mathbf{Y}|\mathbf{v}}(\mathbf{y}) = \frac{1}{(2\pi\sigma_x^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2}\frac{||\mathbf{y} - \alpha\mathbf{v}||^2}{\sigma_x^2}\right), \tag{4.12}$$

où $||\mathbf{y} - \alpha\mathbf{v}||^2 = \sum_{n=0}^{N-1} (y(n) - \alpha v(n))^2$ désigne le carré de la norme du vecteur $\mathbf{y} - \alpha\mathbf{v}$.

Le critère du maximum de vraisemblance est donc équivalent à la recherche de l'état du vecteur $\mathbf{v}$ parmi l'ensemble des états possibles qui minimise la distance euclidienne :

$$D(\mathbf{y}, \mathbf{v}) = ||\mathbf{y} - \alpha\mathbf{v}||^2 = ||\mathbf{y}||^2 + \alpha^2||\mathbf{v}||^2 - 2\alpha\mathbf{y}^t\mathbf{v}. \tag{4.13}$$

Comme $||\mathbf{y}||^2$ ne dépend pas de l'état de $\mathbf{v}$, le critère du maximum de vraisemblance conduit finalement à trouver l'état du vecteur $\mathbf{v}$ qui maximise la métrique :

$$C'(\mathbf{y}, \mathbf{v}) = \mathbf{y}^t\mathbf{v} - \frac{1}{2}\alpha||\mathbf{v}||^2. \tag{4.14}$$

Le terme $\mathbf{y}^t\mathbf{v}$ représente la projection du signal reçu sur l'ensemble des 2^{KL} vecteurs possibles. La complexité de ce traitement est d'autant plus grande que la durée d'observation $N = LN_s$ est longue. Il peut être nettement réduit dès lors que l'Interférence Entre Symboles (IES) est nulle [1]. La détection de chaque K-uplet transmis pendant chaque temps symbole peut alors être effectuée indépendamment des autres. L'observation du signal reçu sur chaque temps symbole ($N = N_s$) suffit. Par la suite et dans tout le développement de la chaîne de tatouage, l'IES sera supposée nulle. La validité de cette hypothèse sera discutée par la suite à la section 7.1.2.

4.3.3 Décision spécifique en fonction de la modulation

La décision quant aux messages numériques possibles dépend de la modulation choisie. On suppose que cette décision est effectuée sur chaque temps symbole.

4.3.3.1 Modulation M-aire

Pour un message numérique sous la forme d'une suite de symboles, le symbole inséré (caractérisant le K-uplet transmis) est directement lié à l'indice du vecteur du dictionnaire choisi pour la modulation. Les valeurs de la variable aléatoire V indiquant l'état du message numérique sur chaque temps symbole sont donc identiques aux valeurs possibles des symboles s_l. La règle de décision consiste alors à rechercher l'indice du dictionnaire de réception qui maximise les métriques de détection $C'(\mathbf{y}, \mathbf{v}) = \mathbf{y}^t\mathbf{v} - \frac{1}{2}\alpha||\mathbf{v}||^2$, avec $\mathbf{v} = \mathbf{d}_m$. Le symbole décidé $\hat{s}_l$ est donc :

$$\hat{s}_l = \arg \max_{m \in [\![0, M-1]\!]} C'(\mathbf{y}, \mathbf{d}_m) \text{ avec } C'(\mathbf{y}, \mathbf{d}_m) = \mathbf{y}^t\mathbf{d}_m - \frac{1}{2}\alpha||\mathbf{d}_m||^2. \tag{4.15}$$

Dans le cas particulier où les vecteurs du dictionnaire d'émission sont de même énergie, c'est à dire que pour tout m la norme $||\mathbf{d}_m||$ est constante, la recherche du maximum des métriques de décision $C'(\mathbf{y}, \mathbf{d}_m)$ peut être simplifiée ; elle devient celle du maximum des métriques de corrélation :

$$\hat{s}_l = \arg \max_{m \in [\![0, M-1]\!]} C(\mathbf{y}, \mathbf{d}_m), \text{ avec } C(\mathbf{y}, \mathbf{d}_m) = \arg \max_{m \in [\![0, M-1]\!]} \mathbf{y}^t\mathbf{d}_m. \tag{4.16}$$

Un tel détecteur est appelé ***détecteur par corrélation***. Il pourra être remarqué que la détection peut être effectuée sans avoir connaissance du facteur d'amplitude α à la réception.

1. L'IES est introduite si les filtres d'émission (incluant la forme des vecteurs du dictionnaire d'émission) et de réception ne respectent pas le critère de Nyquist, ce qui n'est pas le cas pour l'instant.

4.3.3.2 Modulation CDMA

Dans le cas d'une modulation CDMA, le signal modulé $\mathbf{v}$ est une combinaison linéaire des vecteurs du dictionnaire d'émission $\mathcal{D}$ dont les coefficients sont ± 1. Le détecteur optimal consiste donc à détecter le K-uplet vectoriel $\hat{\mathbf{a}}_l$ parmi l'ensemble $\mathfrak{A}$ des K-uplets de composantes ± 1 tel que :

$$\hat{\mathbf{a}}_l = \arg\max_{\mathbf{a}\in\mathfrak{A}} ||\mathbf{y} - \alpha\mathcal{D}\mathbf{a}||^2 \tag{4.17}$$

La détection optimale consiste à examiner l'ensemble des 2^K cas possibles décrits par $\mathfrak{A}$. La complexité d'un tel traitement augmente exponentiellement avec K. Pour diminuer cette complexité, nous pouvons faire appel à un détecteur de complexité linéaire avec K, appelé ***détecteur par décorrélation***. Ce détecteur est optimal si l'énergie consacrée à la transmission de chaque bit est connue du récepteur, mais est sous-optimal si cette énergie ne l'est pas [ZP01]. Il offre en outre l'avantage d'être peu vulnérable à l'interférence entre les vecteurs du dictionnaire [2]. Il traite la projection du signal reçu $\mathbf{y}$ sur les vecteurs du dictionnaire. Le vecteur de corrélation $\check{\mathbf{c}}$ obtenu s'écrit :

$$\check{\mathbf{c}} = \mathcal{D}^t\mathbf{y} = \mathcal{D}^t\left(\alpha\sum_{k=0}^{K-1}(2\mathbf{b}_l - 1)\mathbf{d}_m + \mathbf{x}\right) = \alpha\Gamma_{\mathcal{D}}\mathbf{a}_l + \check{\mathbf{x}}, \tag{4.18}$$

où $\Gamma_{\mathcal{D}} = \mathcal{D}^t\mathcal{D}$ est la matrice d'autocorrélation du dictionnaire d'émission $\mathcal{D}$, $\mathbf{a}_l = 2\mathbf{b}_l - 1$ caractérise le K-uplet émis et $\check{\mathbf{x}} = \mathcal{D}^t\mathbf{x}$ est la projection du signal audio sur les vecteurs du dictionnaire de réception.

L'état du signal $\mathbf{v}$ est entièrement déterminé par l'état du vecteur $\mathbf{a}_l$; celui du signal $\mathbf{y}$ est donné par le vecteur de corrélation $\check{\mathbf{c}}$. Notons $\check{\mathbf{C}}$ le vecteur aléatoire de dimension K modélisant $\check{\mathbf{c}}$. Maximiser la densité de probabilité $p_{\mathbf{Y}|\mathbf{v}}(\mathbf{y})$ sur l'ensemble des messages $\mathbf{v}$ possibles est donc équivalent à maximiser la densité de probabilité de $\check{\mathbf{C}}$ sachant $\mathbf{a}$ parmi l'ensemble des $\mathbf{a}$ possibles. Le signal $\mathbf{x}$ étant supposé BBAG, sa projection sur les K vecteurs du dictionnaire est la réalisation d'un vecteur Gaussien $\check{\mathbf{X}}$. La densité de probabilité de $\check{\mathbf{C}}$ connaissant $\mathbf{a}$ s'écrit alors sous la forme d'une densité de probabilité Gaussienne de dimension K, de moyenne $\Gamma_{\mathcal{D}}$ et de matrice d'autocovariance, celle du bruit $\check{\mathbf{X}}$, de dimension $N_s \times N_s$, notée $\Gamma_{\check{X}}$:

$$p_{\mathbf{C}|\mathbf{a}}(\mathbf{c}) = \frac{1}{(2\pi)^{\frac{K}{2}}\sqrt{\det\left(\Gamma_{\check{X}}\right)}}\exp\left(-\frac{1}{2}\left(\mathbf{c} - \alpha\Gamma_{\mathcal{D}}\mathbf{a}\right)\Gamma_{\check{X}}{}^{-1}\left(\mathbf{c} - \alpha\Gamma_{\mathcal{D}}\mathbf{a}\right)\right), \tag{4.19}$$

où $\det\left(\Gamma_{\check{X}}\right)$ indique le déterminant de la matrice $\Gamma_{\check{X}}$. La règle du maximum de vraisemblance est alors équivalente à trouver le K-uplet $\mathbf{a}$ qui minimise la métrique :

$$\left(\mathbf{c} - \alpha\Gamma_{\mathcal{D}}\mathbf{a}\right)\Gamma_{\check{X}}^{-1}\left(\mathbf{c} - \alpha\Gamma_{\mathcal{D}}\mathbf{a}\right). \tag{4.20}$$

Il s'agit d'un problème d'estimation classique [Hay96], qui peut être résolu par la méthode des moindres carrés. Le minimum est atteint en mettant en œuvre la pseudo-inverse de $\mathcal{D}$; il est donné par $\left(\mathcal{D}^t\mathcal{D}\right)^{-1}\mathcal{D}\mathbf{y}$. Les composantes de $\mathbf{a}$ étant égales à ± 1, on en déduit que :

$$\hat{\mathbf{a}}_l = \operatorname{sign}\left(\alpha\Gamma_{\mathcal{D}}^{-1}\mathcal{D}\mathbf{y}\right). \tag{4.21}$$

2. Cette interférence entre vecteurs est liée aux intercorrélations non nulles existant entre les vecteurs du dictionnaire utilisés pour la modulation.

Le K-uplet binaire reçu est ensuite décidé en fonction du signe des composantes de $\mathbf{a}$: $'0'$ si le signe est négatif et $'1'$ si le signe est positif, soit :

$$\hat{\mathbf{b}}_l = \frac{\hat{\mathbf{a}}_l + 1}{2}. \tag{4.22}$$

Comme dans le cas des modulations M-aire, la détection peut être effectuée sans avoir connaissance de α à la réception. En effet, α étant supposé positif, la décision sur les $\hat{\mathbf{a}}_l$ peut être déduite directement à partir de $\text{sign}\left(\Gamma_{\mathcal{D}}^{-1}\mathcal{D}\mathbf{y}\right)$.

4.3.4 Conséquences sur le choix du dictionnaire d'émission

4.3.4.1 Constellations optimales pour la modulation M-aire

Dans le cas d'une modulation M-aire, nous avons d'ores et déjà établi une première simplification du détecteur (conduisant à un démodulateur par corrélation) lorsque les vecteurs du dictionnaire sont de même norme. Par la suite, nous choisirons donc le dictionnaire d'émission de sorte que :

$$\mathcal{D} = \{\mathbf{d}_m\}_{m\in[\![0,M-1]\!]}\,, \text{ avec } \forall m \in [\![0, M-1]\!], \frac{1}{N_s}\mathbf{d}_m^t\mathbf{d}_m = 1, \tag{4.23}$$

l'énergie consacrée à la transmission de chaque symbole étant réglée par le facteur α.

Sous cette hypothèse, le détecteur optimal consiste à rechercher le vecteur du dictionnaire dont la corrélation avec le signal reçu est maximale. La détection peut donc être effectuée dans l'espace engendré par les vecteurs du dictionnaire. La constellation engendrée par ces vecteurs influe sur le résultat de cette détection et donc sur la probabilité d'erreur de transmission. La théorie des communications numériques [Pro01] montre que la probabilité d'erreur de transmission est d'autant plus petite que la distance entre les vecteurs est grande. En effet, dans le cas d'un dictionnaire de deux vecteurs opposés $\mathbf{d}_1 = -\mathbf{d}_0$ et de puissance unité, la probabilité d'erreur de transmission s'exprime sous la forme :

$$P_e = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}}\right) = \frac{1}{2}\text{erfc}\left(\frac{\alpha d}{2\sqrt{2}\sigma_x}\right), \tag{4.24}$$

où σ_x^2 est la variance du bruit et $d = ||\mathbf{d}_1 - \mathbf{d}_0|| = 2||\mathbf{d}_0|| = 2\sqrt{N_s}$ est la distance entre les deux vecteurs. Pour maximiser cette corrélation, deux constellations peuvent être envisagées :

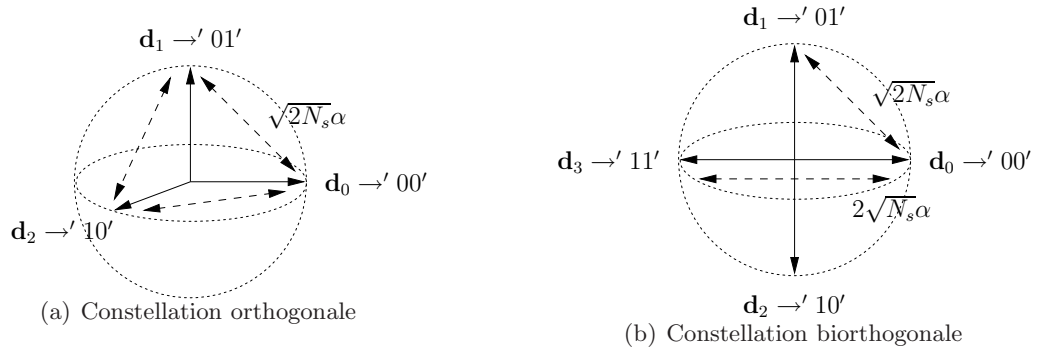
- une ***constellation orthogonale*** ; le dictionnaire contient M vecteurs orthonormés définissant un espace de dimension M :

$$\mathcal{D} = \{\mathbf{d}_m\}_{m\in[\![0,M-1]\!]}\,, \text{ avec } \forall (m,p) \in [\![0, M-1]\!]^2, \frac{1}{N_s}\mathbf{d}_m^t\mathbf{d}_p = \delta(m,p), \tag{4.25}$$

  où $\delta(m,p)$ est le symbole de Kronecker (égal à 1 si $m = p$ et à 0 sinon).

- une ***constellation bi-orthogonale*** ; le dictionnaire contient M vecteurs répartis en deux sous-ensembles. Le premier sous-ensemble contient $M/2$ vecteurs orthonormés, le second contient les vecteurs opposés du premier. Le dictionnaire définit donc un espace de dimension $M/2$. Pour minimiser le nombre d'erreurs de transmission, le dictionnaire est indexé de façon à permettre un encodage source de Gray [Pro01]. Il s'écrit :

$$\mathcal{D} = \{\mathbf{d}_m\}_{m\in[\![0,M-1]\!]}\,, \text{ avec } \begin{cases} \forall (m,p) \in [\![0, \frac{M}{2}-1]\!]^2, \frac{1}{N_s}\mathbf{d}_m^t\mathbf{d}_p = \delta(m,p), \\ \forall m \in [\![0, \frac{M}{2}-1]\!], \mathbf{d}_{M-1-m} = -\mathbf{d}_m. \end{cases} \tag{4.26}$$



(a) Constellation orthogonale

(b) Constellation biorthogonale

Figure 4.2 – Représentation des constellations possibles d'un dictionnaire $\mathcal{D} = \{\mathbf{d}_0, \mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ de $M = 4$ vecteurs, pour lequel $\mathbf{d}_0$ encode le uplet $'00'$, $\mathbf{d}_1$ le uplet $'01'$, $\mathbf{d}_2$ le uplet $'10'$ et $\mathbf{d}_3$ le uplet $'11'$: (a) constellation orthogonale (le 4ème vecteur n'est ici pas représenté), (b) constellation biorthogonale.

La figure 4.2 propose une représentation de ces deux constellations dans le cas où $M = 4$.

4.3.4.2 Constellations optimales pour la modulation CDMA

Dans le cas d'une modulation CDMA, la détection sous-optimale par décorrélation a pour but de supprimer l'interférence existant entre chaque vecteur porteur d'une des composantes du K-uplet transmis. En effet, l'inverse de la matrice d'autocorrélation $\Gamma_{\mathcal{D}}$ du dictionnaire d'émission compense les corrélations non nulles existant entre les vecteurs. Pour faciliter cette décorrélation, il convient de choisir des vecteurs mutuellement orthogonaux (puisque induisant une matrice d'autocorrélation proportionnelle à la matrice identité) et de même norme (puisque consacrant la même énergie à la transmission de chaque bit). Le dictionnaire d'émission est donc choisi pour contenir M vecteurs orthogonaux de puissance unité :

$$\mathcal{D} = \{\mathbf{d}_m\}_{m \in [\![0, M-1]\!]}, \text{ avec } \forall (m,p) \in [\![0, M-1]\!]^2, \frac{1}{N_s}\mathbf{d}_m^t \mathbf{d}_p = \delta(m,p). \tag{4.27}$$

D'une manière plus générale, pour de nombreuses applications CDMA (notamment celles de transmissions asynchrones), les propriétés d'intercorrélation sont aussi importantes que les propriétés d'autocorrélation. Ces applications recourent à des séquences pseudo-aléatoires spécifiquement construites pour leur propriétés d'intercorrélation telles que les séquences-m [Gol81], les séquences Gold [McE87] ou les séquences Kasami [Pro01]. Dans un contexte de tatouage, nous avons montré [Bar02] que ces séquences ne permettent pas d'apporter d'améliorations majeures quant aux performances de la chaîne de tatouage : les taux d'erreurs binaires obtenus sont identiques que l'on utilise des séquences pseudo-aléatoires spécifiques ou des vecteurs aléatoires construits par simple réalisation d'un bruit blanc et orthonormalisation. De plus, ces séquences présentent l'inconvénient d'imposer au débit de transmission des valeurs prédéfinies (dépendant du nombre de registres à décalage utilisés pour leur construction). Le débit de fonctionnement de la chaîne de tatouage devient moins modulable. De fait, ces séquences ne seront pas considérées par la suite.

4.4 Performances théoriques

La théorie des communications numériques [Pro01] établit les performances, en terme de probabilité d'erreur binaire P_e, d'une chaîne de tatouage. Dans cette section, nous adaptons les résultats théoriques et nous présentons les courbes de performances théoriques qui en découlent au regard des conditions de transmission du système de tatouage qui seront établies par la suite.

4.4.1 Cas des modulations M-aire

Dans le cas des modulations M-aire, qui réalisent la transmission d'un K-uplet binaire sous la forme d'un symbole à valeur dans l'alphabet $\{0, \ldots, 2^K - 1\}$, il convient de distinguer la probabilité d'erreur binaire, probabilité d'erreur sur les bits reçus notée P_e, de la probabilité d'erreur sur les symboles reçus, notée P_s. Ces deux probabilités sont liées d'après [Pro01] par la relation [3] :

$$P_e = \frac{1}{2}\frac{M}{M-1}P_s. \tag{4.28}$$

Cette relation induit deux remarques :

- Dans un premier temps, la probabilité d'erreur binaire est toujours inférieure à la probabilité d'erreur symbole. En effet, un symbole erroné $s_l = \sum_{k=0}^{K-1} b_l[k]2^k$ n'entraîne une erreur sur tous les bits que si le symbole émis est $s'_l = \sum_{k=0}^{K-1} b_l\bar{[}k]2^k$ avec $b_l\bar{[}k] = 0$ si $b_l = 1$ et $b_l\bar{[}k] = 1$ si $b_l = 0$.
- D'autre part, la probabilité d'erreur binaire décroît lorsque M augmente. D'ailleurs, lorsque M tend vers l'infini, $P_e \approx \frac{P_s}{2}$, signifiant que la répartition des erreurs binaires dans chaque symbole erroné tend vers une répartition "équiprobable", c'est à dire un bit erroné sur deux.

L'expression de la probabilité d'erreur dépend de la constellation choisie pour le dictionnaire d'émission :

- dans le cas d'un dictionnaire orthogonal, elle prend la forme suivante :

$$\begin{aligned} P_e &= \frac{1}{2}\frac{M}{M-1}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(1-\left[1-\frac{1}{2}\mathrm{erfc}\left(\frac{\tau}{\sqrt{2}}\right)\right]^{M-1}\right)\ldots \\ &\ldots\exp\left(-\frac{1}{2}\left(\tau-\sqrt{\frac{\alpha^2 N_s}{\sigma_x^2}}\right)^2\right)d\tau. \end{aligned} \tag{4.29}$$

3. En effet [Pro01, Lar05], à supposer des symboles émis équiprobables, une erreur sur un symbole se produit avec une probabilité $\frac{P_s}{M-1}$. Cette erreur sur le symbole se répercute sur les K bits qui le constituent et peut entraîner k erreurs binaires avec k allant de 1 à K. De plus, la répartition de ces k bits erronées (dans le M-uplet formé par le symbole) suit l'une des $\binom{K}{k} = \frac{K!}{k!(K-k!)}$ répartitions possibles, où $k!$ désigne la factorielle de k. Le nombre moyen de bit erroné par symbole est donc $N_{be} = \sum_{k=1}^{K}\binom{K}{k}\frac{P_s}{M-1} = K\frac{2^{K-1}}{2^K-1}P_s$. La probabilité d'erreur moyenne par bit est alors déduite par $P_b = \frac{N_{be}}{K}$.

- dans le cas d'un dictionnaire biorthogonal, elle devient [4] :

$$P_e = \frac{1}{2}\frac{M}{M-1}\frac{1}{\sqrt{2\pi}}\int_0^{\infty}\left(1-\left[\frac{1}{\sqrt{2\pi}}\int_{-\tau}^{\tau} exp\left(-\frac{\theta^2}{2}\right)d\theta\right]^{\frac{M}{2}-1}\right)\ldots$$
$$\ldots\exp\left(-\frac{1}{2}\left(\tau-\sqrt{\frac{\alpha^2 N_s}{\sigma_x^2}}\right)^2\right)d\tau. \tag{4.30}$$

Le rapport $\frac{\alpha^2 N_s}{\sigma_x^2}$ est le rapport entre l'énergie du signal de tatouage $\mathbf{t}$ consacrée à la transmission d'un symbole et la puissance du bruit audio $\mathbf{x}$. Chaque symbole encodant $K = \log_2(M)$ bits, notons $N_b = \frac{N_s}{K}$ le nombre d'échantillons consacrés à la transmission d'un bit. Il est alors possible de réécrire les expressions de P_e en fonction du rapport $\frac{\alpha^2 N_b}{\sigma_x^2}$.

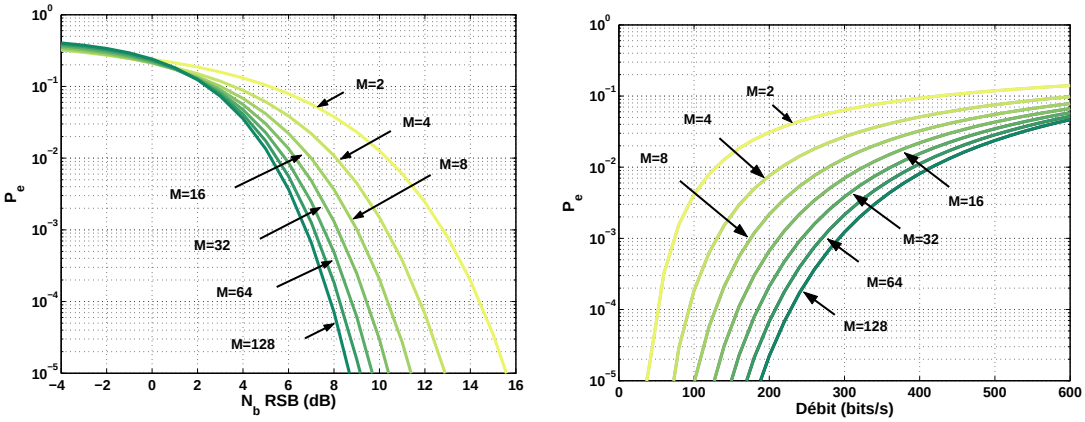
Le rapport $\frac{\alpha^2 N_b}{\sigma_x^2}$ est homogène au paramètre ***ebno*** utilisé par les théoriciens des communications numériques pour caractériser les performances d'une chaîne de communication. Ce paramètre traduit les conditions de transmission de chaque bit indépendamment du temps utilisé pour la transmission [5]. De fait, les performances d'une chaîne de tatouage (y compris dans la littérature) sont souvent établies en mesurant la probabilité d'erreur en fonction de l'ebno. De plus, le rapport $\frac{\alpha^2}{\sigma_x^2}$ est homogène à un RSB - le RSB caractérisant la puissance du signal porteur de l'information $\mathbf{t}$ par rapport à celle du bruit audio $\mathbf{x}$ en entrée du corrélateur. L'étude des performances de la chaîne de tatouage qui sera menée section 7.1 montre que ce rapport varie, en amont du corrélateur, entre -20 et 25 dB [6]. L'expression de la probabilité d'erreur peut donc être établie en fonction de ce RSB et du débit de transmission binaire $R = F_e/N_b$ en substituant à N_s l'expression $\log_2(M)\frac{F_e}{R}$.

Nous présentons figure 4.3 et 4.4 les probabilités d'erreur P_e obtenues en fonction de M dans le cas d'un dictionnaire orthogonal puis biorthogonal. Les figures 4.3 (a) et 4.4 (a) proposent, pour chaque configuration de dictionnaire, les variations de P_e en fonction de $N_b RSB = \frac{\alpha N_b}{\sigma_x^2}$. Les figures 4.3 (b) et 4.4 (b) présentent, elles, les variations de P_e en fonction du débit R et de la constellation du dictionnaire dans une configuration où le RSB est peu favorable (ici un RSB imposé de -15 dB). Ces courbes mettent en évidence l'apport de l'augmentation de la taille du dictionnaire sur la probabilité d'erreur et l'existence de valeurs limites pour N_bRSB. En effet, quelle que soit la constellation engendrée par les vecteurs du dictionnaire, en augmenter le nombre améliore la probabilité d'erreur à condition que le paramètre N_bRSB soit supérieur à une valeur limite. Au delà de cette valeur, la tendance s'inverse. Cette limite est de l'ordre de 0 dB dans le cas d'un dictionnaire orthogonal et de l'ordre de 2 dB dans le cas d'un dictionnaire

4. Cette expression se simplifie aisément lorsque le dictionnaire est constitué seulement de deux vecteurs antipodaux (cas où $M = 2$), donnant lieu à l'expression classique de la probabilité d'erreur d'une modulation binaire antipodale, déjà rappelée équation 4.24.

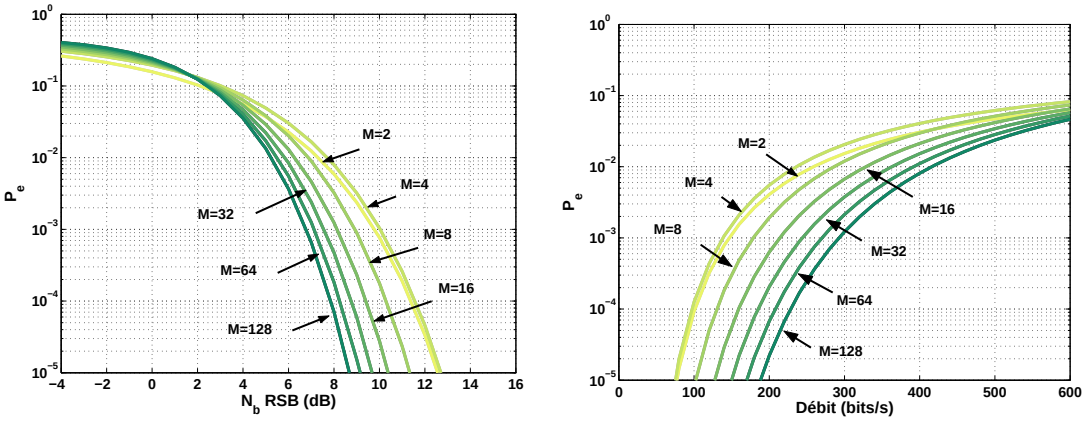
5. Le paramètre ebno est le rapport entre l'énergie conférée à la transmission d'un bit pendant le temps bit N_b noté par convention E_b et la puissance d'un bruit BBAG de moyenne nulle et de variance $\sigma_n^2 = \frac{N_0}{2}$.

6. Le RSB est notamment dépendant de l'efficacité des filtres introduits pour les besoins de la chaîne de tatouage et des conditions de débit.



(a) (b)

Figure 4.3 – Probabilité d'erreur binaire des modulations M-aire pour différentes tailles M d'un dictionnaire orthogonal en fonction : (a) du paramètre N_bRSB et (b) du débit R pour un RSB fixé à -15 dB.



(a) (b)

Figure 4.4 – Probabilité d'erreur binaire des modulations M-aire pour différentes tailles M d'un dictionnaire biorthogonal en fonction : (a) du paramètre N_bRSB et (b) du débit R pour un RSB fixé à -15 dB.

biorthogonal (pour $M > 2$)[7]. Étant donnés les débits de transmission choisis (inférieurs à 1 kbit/s) et l'échelle de valeurs du RSB, la valeur limite de N_bRSB sera rarement atteinte par notre système de tatouage. Ainsi dans le cas d'un dictionnaire orthogonal, même pour un RSB relativement peu favorable (ici -15 dB), la fiabilité de la transmission à débit fixé croît avec M, ce que confirme la figure 4.3 (b). Dans le cas d'un dictionnaire biorthogonal, la limite étant plus élevée, elle est atteinte aux alentours de 400 bit/s pour $M = 8$ vecteurs : en effet, la figure 4.4 (b) montrent que les probabilités d'erreur obtenues dans le cas binaire antipodal ($M = 2$) deviennent

7. Ces valeurs sont données à titre d'ordre de grandeur ; la précision de la visualisation ne permet pas d'avoir ici une estimation fiable de ces valeurs limites.

meilleures que celle de $M = 8$. Une attention particulière doit également être portée au cas d'une modulation bi-orthogonale avec $M = 4$: les probabilités d'erreur obtenues dans cette configuration sont toujours inférieures à celles d'une modulation binaire antipodale.

4.4.2 Cas des modulations CDMA

Les modulations CDMA ont été conçues pour acheminer dans un même canal plusieurs messages numériques envoyés parallèlement sur chacun des vecteurs du dictionnaire : le but de cette modulation n'est donc pas d'améliorer la probabilité d'erreur d'un système de transmission mais bien d'augmenter la "capacité" de transmission du système. Chaque temps symbole N_s permet la transmission simultanée des K bits du K-uplet binaire $\mathbf{b}_l$ et conduit au signal reçu :

$$\mathbf{y} = \mathbf{x} + \mathbf{t} = \frac{\alpha}{\sqrt{K}}(2b_l[0] - 1)\mathbf{d}_0 + ... + \frac{\alpha}{\sqrt{K}}(2b_l[K-1] - 1)\mathbf{d}_{K-1} + \mathbf{x}, \tag{4.31}$$

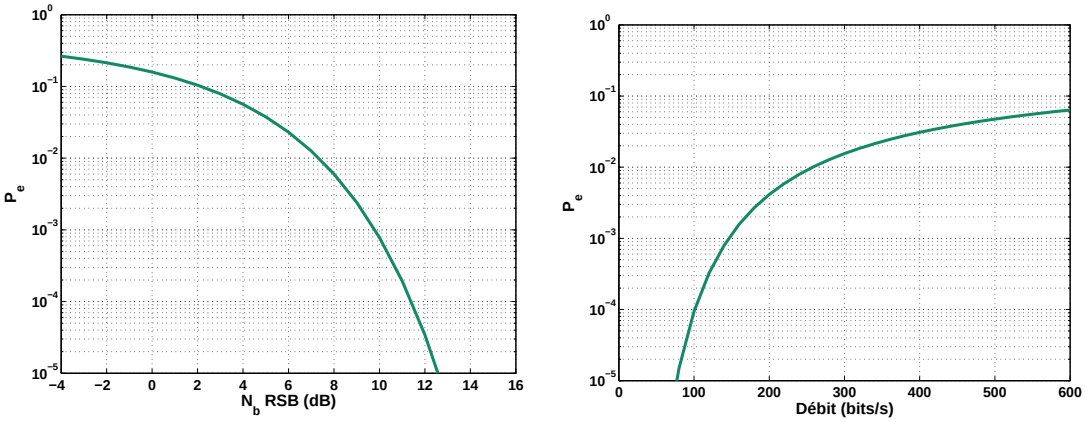
le facteur $\frac{1}{\sqrt{K}}$ étant choisi pour limiter la puissance du signal modulé (imposé à 1 dans le contexte du tatouage). Transmettre le bit $b_l[0]$ peut donc être vu comme le résultat d'une modulation binaire antipodale dans une chaîne de communication où le bruit provient de la transmission des autres bits du K-uplet binaire et du bruit audio $\mathbf{x}$. Le récepteur projetant le signal reçu sur les vecteurs du dictionnaire, alors, si ces vecteurs sont orthogonaux, la transmission des bits autres que $b_l[0]$ n'interfère pas sur la détection de $b_l[0]$. Dans ce cas particulier, la modulation CDMA est équivalente à K chaînes de transmission parallèles bruitées par le même bruit $\mathbf{r}$. La probabilité d'erreur qui en est déduite est donc celle d'une modulation binaire antipodale (pour laquelle un bit est transmis avec une énergie $\frac{1}{K}$ pendant une durée N_s) :

$$P_e = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{\alpha^2 N_s}{2K\sigma_x^2}}\right). \tag{4.32}$$

Comme dans le cas des modulations M-aire, le rapport $\frac{\alpha^2}{\sigma_x^2}$ est homogène au RSB, tandis que $N_b = \frac{N_s}{K}$ est équivalent à la durée en échantillon pour transmettre un bit et s'obtient directement en fonction du débit de transmission par $N_b = \frac{F_e}{R}$. Il s'en déduit, qu'à débit de transmission fixé, la probabilité d'erreur du système est indépendante du nombre K de bits transmis simultanément (sous réserve que les vecteurs du dictionnaire de réception soient orthogonaux et de même puissance). Ses variations en fonction du débit de transmission coïncident donc avec les performances de la modulation M-aire dans le cas d'un dictionnaire de 2 vecteurs antipodaux ($K = 1$, $M = 2$). Nous rappelons néanmoins figure 4.5 l'évolution de cette probabilité d'erreur en fonction du paramètre N_bRSB et en fonction du débit lorsque le RSB est peu favorable (ici -15 dB).

Conclusion

Dans ce chapitre, nous avons rappelé les outils de la théorie des communications numériques établissant une chaîne de communication optimale pour un bruit de canal blanc Gaussien.



(a) (b)

Figure 4.5 – Probabilité d'erreur binaire des modulations CDMA quelle que soit la taille du dictionnaire en fonction en fonction du paramètre E_bRSB pour la figure (a) et du débit R pour un RSB fixé à -15 dB pour la figure (b).

Ces outils peuvent être exploités pour la construction du système de tatouage, qui fait l'objet du chapitre suivant. Cette chaîne se caractérise néanmoins par des contraintes très particulières nécessitant d'adapter ces outils théoriques. D'une part, le signal de tatouage doit être inaudible en présence du signal audio. Cette inaudibilité ne peut être obtenue par le choix d'un simple facteur d'amplitude régissant le rapport de puissance entre le signal de tatouage et le bruit audio. D'autre part, le signal audio, bruit de la chaîne de tatouage, est très différent d'un bruit blanc additif gaussien. Le récepteur optimal défini par la théorie des communications numériques doit donc être modifié pour prendre en compte les caractéristiques du bruit audio (un bruit fortement non-stationnaire et coloré).

Chapitre 5

Construction d'un système de tatouage audio

Construire une chaîne de tatouage en faisant l'hypothèse d'une insertion dans le domaine temporel fait appel aux outils de la théorie de communications numériques que nous avons présentés dans le chapitre précédent. Malheureusement ces outils s'attachent à un modèle théorique dans lequel le signal de tatouage serait le résultat d'une simple modulation et le bruit audio serait un bruit blanc additif Gaussien. Cette modélisation est très éloignée des contraintes de transmission spécifiques de l'application de tatouage. Ces contraintes, à savoir l'inaudibilité et la fiabilité de la transmission, sont toutes deux liées aux propriétés du signal audio. Elles requièrent de modifier de manière spécifique les organes de la chaîne de communication théorique développée précédemment.

Garantir l'inaudibilité du tatouage ne peut être atteint par une simple modulation paramétrée par un facteur d'amplitude α qui conduirait à un tatouage de faible puissance devant le signal audio. Cette contrainte requiert la prise en compte d'éléments de la psychoacoustique, pour définir les conditions d'inaudibilité exactes à satisfaire par le tatouage et modifier en conséquence la structure de l'émetteur. Les modifications de l'émetteur qui en sont déduites sont présentées section 5.1.

Garantir la transmission fiable de l'information nécessite de prendre en compte la statistique du signal audio. Celle-ci est très différente de celle d'un canal BBAG étudié par la théorie. En effet, les signaux audio sont notamment connus pour être non-stationnaires et fortement colorés. Nous ferons l'hypothèse que le signal audio peut être modélisé [1] par un processus Auto-Régressif (AR) d'ordre P_0 globalement non stationnaire mais supposé localement stationnaire. Au regard

1. Cette modélisation apparaît couramment dans les travaux sur la reconnaissance et la synthèse sonore, tant sur la production des signaux de parole [RJ93] que sur la modélisation des instruments. Ce modèle coïncide notamment avec le modèle d'excitation-résonance ou source-filtre [Zöl03]. Le son est produit en filtrant un signal d'excitation (souvent vu comme un bruit blanc) par un filtre de synthèse. Ce filtre modélise les résonances du signal et peut être formalisé par un filtre tout-pôle $A(f)$ dont la réponse en fréquence coïncide avec la densité spectrale de puissance du signal audio $S_x(f)$:

$$A(f) = 1/\sum_{k=0}^{P_0} a_k e^{-j2\pi f},$$

où P_0 est l'ordre du filtre. Ce filtre confère au signal audio une structure auto-régressive. Les coefficients de ce filtre devront être actualisés sur chaque fenêtre d'analyse pour laquelle le signal audio est localement stationnaire.

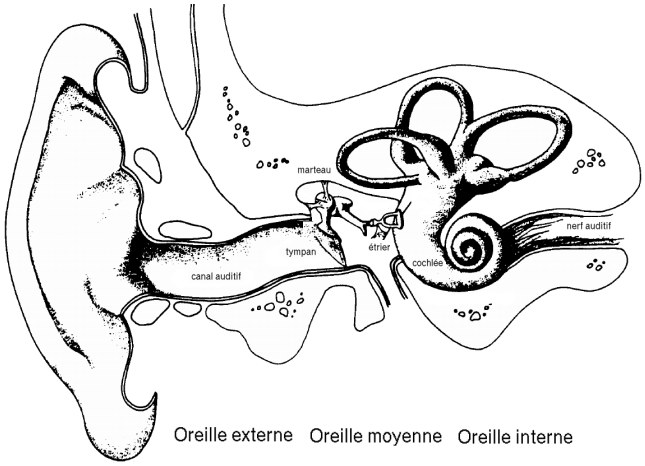


Figure 5.1 – Vue générale de l'oreille (reproduit d'après [Goy02]).

de cette modélisation, le récepteur optimal défini pour un canal BBAG sera généralisé au cas d'un bruit AR. Ce récepteur devra également prendre en compte les choix spécifiques réalisés par l'émetteur pour construire le tatouage inaudible $t(n)$. Les détails de sa conception sont présentés section 5.2.

Dans les développements à venir, nous supposerons l'IES nulle. Cette dernière fera l'objet d'une étude spécifique réalisée section 7.1.2 et incorporant l'ensemble des éléments du système de tatouage.

5.1 Modulation

La seule contrainte imposée au signal de tatouage $t(n)$ et au signal modulé $v(n)$ dont il est issu est la contrainte d'inaudibilité. Il est nécessaire de faire appel à la psychoacoustique pour établir les conditions d'une transmission inaudible et choisir l'allure des formes d'onde du dictionnaire à partir desquelles est construit $v(n)$.

5.1.1 Approcher l'inaudibilité par l'utilisation d'un modèle psychoacoustique

5.1.1.1 Quelques éléments de psychoacoustique

La psychoacoustique s'attache à modéliser le système auditif humain et la perception du son. Ses résultats, essentiellement basés sur des expérimentations, mettent en évidence les défauts du système auditif, et plus particulièrement les phénomènes de masquage. Le masquage est le procédé par lequel un son, appelé ***son masqué*** est rendu inaudible en présence d'un autre, appelé ***son masquant***. Ce phénomène, particulièrement exploité en compression audio, peut également être appliqué au tatouage pour déterminer les conditions de masquage du signal modulé $v(n)$ en présence du signal audio $x(n)$.

Anatomie de l'oreille et perception du son

La perception d'un son met en œuvre des procédés acoustiques, mécaniques et nerveux, qui convertissent des variations de pression acoustique en un message nerveux à destination du cerveau [Fil04]. L'oreille est divisée en trois éléments appelés ***oreille externe***, ***oreille moyenne*** et ***oreille interne*** et présenté schématiquement figure 5.1. L'onde sonore, caractérisée par un niveau d'intensité exprimé en décibel Sound Pression Level (SPL), est captée par l'oreille externe, où elle est amplifiée sélectivement en fréquence avant de stimuler le tympan. Cette stimulation provoque la vibration des osselets de l'oreille moyenne et l'adaptation d'impédance entre le milieu aérien de l'oreille externe et le milieu liquidien de l'oreille interne. Elle est ensuite traduite en un message nerveux par l'organe d'audition de l'oreille interne : la cochlée. La cochlée est reliée à la membrane basilaire sur laquelle repose l'organe de Corti, constitué d'un ensemble de cellules cilliées. La résonance de la membrane basilaire en réponse à l'onde sonore provoque l'excitation de ces cellules cilliées, qui réalisent la conversion du signal sonore en message électrochimique à destination des nerfs auditifs. Étant donnée la structure anatomique particulière de la membrane basilaire (qui devient d'autant plus large et d'autant plus flexible que l'on se déplace de la base de la cochlée à son extrémité), le maximum de résonance est atteint en un point spécifique de la membrane, fonction de la tonalité de l'onde sonore. Une concordance entre les points de la membrane et l'échelle des fréquences audibles peut être établie. L'oreille peut donc être vue comme un analyseur fréquentiel du signal sonore.

Seuil d'audition absolu

L'étendue des puissances acoustiques susceptibles de stimuler le nerf auditif sans douleur est vaste, allant de 0 à 120 dB SPL. Néanmoins, un son n'est détecté (et ne produit une résonance significative de la membrane basilaire) que si son niveau d'intensité est supérieure à un seuil fonction de la fréquence. Les expériences de Fletcher ont permis de quantifier ce seuil (sous la forme d'une quantité d'énergie moyenne en dB SPL) pour un son sinusoïdal pur de fréquence variable dans un environnement sonore non bruité. Cette quantification prend la forme d'une courbe appelée ***seuil d'audition absolu*** qui peut être approchée par la fonction non linéaire :

$$S_a(f) = 3,64 f^{-0,8} - 6,5 \exp\left(-0,6(f-3,3)^2\right) + 10^{-3} f^4 \text{ (dB SPL)}, \tag{5.1}$$

où f désigne la fréquence en kHz.

Bandes critiques

L'analyse fréquentielle réalisée par le système auditif peut être modélisée par un banc de filtres passe-bande, appelés ***filtres auditifs***, dont les bandes passantes, appelées ***bandes critiques***, recouvrent en se chevauchant la gamme des fréquences audibles (de 20 Hz à 20 kHz). Ce banc de filtres caractérise efficacement le mode de perception de l'oreille, son pouvoir de séparabilité et sa résolution fréquentielle.

Ce modèle s'est dégagé à la suite de différentes expériences (réalisées notamment par Fletcher en 1940 puis par Greenwood et Zwicker vers les années 60) sur la perception de l'intensité sonore [PS00] et sur les phénomènes de masquage. Citons plus particulièrement deux d'entre elles :

- La première a trait au seuil d'audition absolu. Considérons une première sinusoïde de fréquence f_1 et de puissance σ_x^2 suffisamment forte pour être détectée par l'oreille (c'est à dire de puissance plus grande que le seuil d'audition absolu S_a en f_1). Ajoutons une deuxième sinusoïde de fréquence f_2 proche de f_1 et de même puissance σ_x^2. Les mesures

expérimentales montrent que l'ensemble des deux sinusoïdes est audible si $2\sigma_x^2 \geq S_a(f_1)$. La puissance nécessaire pour détecter les deux sinusoïdes est donc plus faible que dans le cas d'une sinusoïde seule. Cette propriété reste valable tant que l'écart entre les deux fréquences n'excèdent pas une valeur seuil, correspondant à la bande critique.

- La seconde expérience s'intéresse à l'intensité perçue d'un son. Elle montre que l'intensité perçue (ou ***sonie***) d'un bruit à bande étroite de niveau d'intensité en dB SPL reste constante tant que la largeur de bande du bruit ne dépasse pas la largeur de bande critique. Au delà, l'intensité perçue augmente.

Ces expériences, aux objectifs différents, ont conduit à la même notion de bandes critiques et ont révélé approximativement les mêmes largeurs de bande : ces largeurs sont non-uniformes et augmentent avec la fréquence. Elles permettent également d'admettre que la puissance perçue par l'oreille dans une bande critique est égale à la somme de toutes les puissances des composantes dans cette bande de fréquence : l'oreille réalise une intégration des puissances dans une bande critique.

Une nouvelle échelle des fréquences, celle des Bark, a donc été définie pour faciliter les traitements dans les bandes critiques. Fréquences en kHz (f_H) et fréquences en Bark (f_B) sont reliées par :

$$f_B = 13 \arctan\left[0,76 f_H\right] + 3,5 \arctan\left[\left(\frac{f}{7,5}\right)^2\right]. \tag{5.2}$$

Masquage

Le masquage est le phénomène par lequel un son, le ***son masqué***, peut être rendu inaudible par la présence d'un second, le ***son masquant***. Deux types de masquage peuvent être distingués : le masquage temporel, lorsque les deux sons se produisent à intervalle de temps rapproché mais disjoint, et le masquage fréquentiel ou simultané, lorsque les deux sons se produisent simultanément.

- Le masquage temporel peut être de deux natures, en fonction de l'ordre d'apparition des sons masquant et masqué. Si le son masqué est antérieur au son masquant, on parle de ***pré-masquage*** ; sinon, de ***post-masquage***. Un son n'est masqué que si son intensité est inférieure à un certain seuil. Ce seuil est fonction de l'écart temporel entre les deux sons [2], de la fréquence des deux sons, et de l'intensité et de la durée du son masquant. Il est illustré figure 5.2 en fonction du temps.

- Le masquage simultané dépend des spectres d'amplitude du son masquant et du son masqué. Là encore, le son est masqué si son intensité est inférieure à un seuil, fonction de l'intensité du masquant, des fréquences respectives des sons masqué et masquant et de leurs natures (sinusoidales ou bruitées). Ce phénomène peut être expliqué d'un point de vue anatomique. La présence du son masquant de forte intensité, provoque la vibration de la membrane basilaire. Cette vibration est relativement forte au point spécifique de la membrane correspondant à la fréquence du masquant et atténuée dans les zones avoisinantes. Une seconde vibration, liée au son masqué, ne peut donc être détectée que si elle dépasse la vibration (même atténuée) provoquée par le son masquant.

2. Cet écart est en fait l'écart d'apparition entre les sons dans le cas du pré-masquage et de l'écart entre la disparition du masquant et de l'apparition du masqué dans le cas du post-masquage

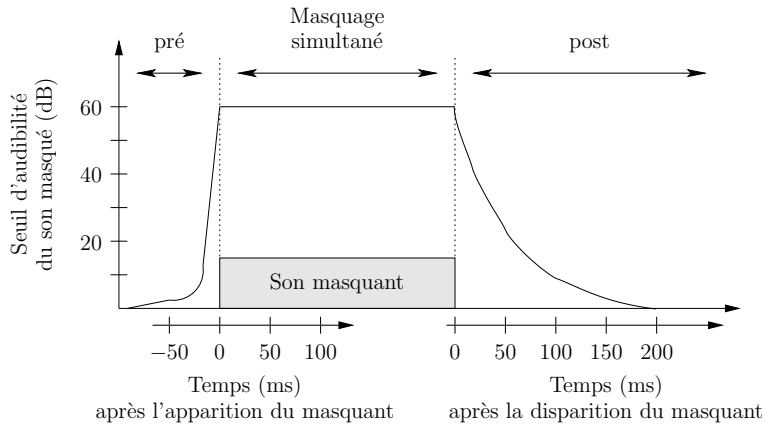


Figure 5.2 – Seuil d'inaudibilité du son masqué en fonction de l'ordre et du temps d'apparition du son masqué et du son masquant.

La nature des sons, sinusoidaux ou bruités, jouent un rôle important dans les caractéristiques du phénomène de masquage. Quatre cas de figure, correspondant aux 4 combinaisons sinus-bruit qui peuvent être choisies pour le son masqué et le masquant, ont été plus particulièrement analysés. Ces études ont permis de représenter l'effet de masquage sous la forme d'une courbe, appelée ***fonction d'étalement***, d'allure triangulaire et dont les paramètres dépendent de la nature des sons masquants et masqués. La figure 5.3 présente l'allure de ces courbes dans le contexte d'étude suivant : le son masquant est un bruit bande étroite centré sur la fréquence $f_M = 1$ kHz et dont la puissance σ_M^2 est variable. Le son masqué est une sinusoïde de fréquence f_m et de puissance σ_m^2. Les courbes présentent la valeur maximale de σ_m^2 en fonction de la fréquence f_m pour laquelle le son masqué est inaudible et ce pour différentes valeurs de σ_M^2. On constate que les fonctions d'étalement coïncident avec le seuil d'audition absolu lorsque f_m est relativement éloignée de f_M, tandis qu'elles prennent une forme triangulaire lorsque f_m est proche de f_M. Les fonctions d'étalement peuvent donc être modélisées par la seule donnée de trois paramètres : l'***indice de masquage*** qui correspond au rapport de puissance σ_M^2/σ_m^2 lorsque f_m et f_M coïncident, la pente gauche et la pente droite [3]. Ces trois paramètres dépendent eux aussi de façon fortement non linéaire de la fréquence, de l'intensité et de la nature du son masquant et du masqué [4].

5.1.1.2 Modèle psychoacoustique et seuils de masquage

La théorie de la psychoacoustique donne les conditions à satisfaire par le signal de tatouage dans le cas idéal où le signal audio serait une sinusoïde pure ou un bruit à bande étroite. Ces conditions prennent la forme d'une fonction d'étalement qui indique la puissance limite du signal de tatouage en fonction de sa fréquence. Le signal audio est bien entendu beaucoup plus complexe qu'une simple sinusoïde. Il faudrait pouvoir combiner les fonctions d'étalement de chaque composante fréquentielle du signal pour établir une courbe, appelée ***seuil de masquage***,

3. si les fréquences sont exprimées en Bark et les puissances en dB.

4. Rappelons à titre d'illustration que l'indice de masquage est de l'ordre de 4 dB lorsque le son masquant est un bruit à bande étroite et le son masqué une sinusoïde, alors qu'il est d'environ 24 dB pour la configuration inverse [PS00].

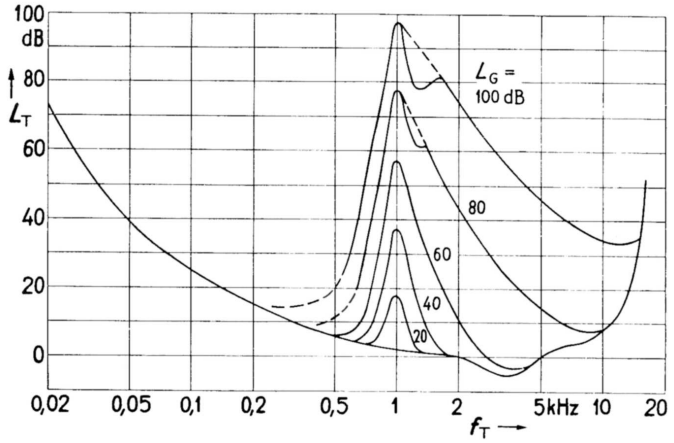


Figure 5.3 – Fonctions d'étalement pour des bruits à bande étroite. Le signal masquant est un bruit de bande centré sur 1 kHz de largeur 160 Hz présenté aux niveaux L_G = 100, 80, 60, 40 et 20 dB. D'après [ZF81].

caractérisant le pouvoir masquant du signal audio sur toute sa gamme de fréquence. La difficulté de cette transformation tient au fait que la théorie de la psychoacoustique n'établit pas de quelle façon ses fonctions d'étalement individuelles peuvent être combinées.

Ce problème s'est déjà posé pour les applications de compression audio : établir un seuil de masquage permet de déterminer l'allure fréquentielle maximale tolérée du bruit introduit par la quantification du signal audio. Deux ***Modèles PsychoAcoustiques*** (MPA) adaptés au contexte de la compression audio ont donc été proposés. Ces deux modèles, le modèle n°1 et le modèle n°2, diffèrent par leur gestion des indices de masquage associés aux fonctions d'étalement [5]. Construits expérimentalement, ils ont par la suite été validés par des tests d'écoute [Nor93].

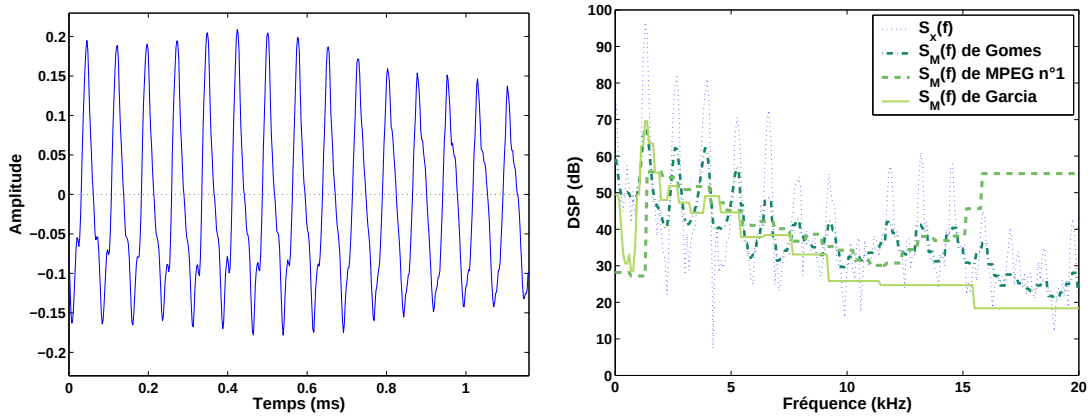
Une telle étude n'a pas encore été réalisée pour les applications de tatouage. La majorité des systèmes basés sur des modèles psychoacoustiques [BTH96, LGP04] ont recours aux modèles définis pour MPEG, dont la mise en œuvre dans un contexte de tatouage est contestable. Seuls deux systèmes ont proposé un seuil de masquage dans un contexte de tatouage : celui de Gomes développé dans [Gom02] pour un tatouage additif dans le domaine temporel et celui de Garcia détaillé dans [Gar99] pour un tatouage substitutif dans le domaine fréquentiel. La figure 5.4 présente à titre indicatif ces trois seuils de masquage calculés pour un extrait de violon [6]. Dans les sections suivantes, nous détaillons les principes de mise en œuvre de chacun de ces seuils. Le choix de l'un d'entre eux pour notre système de tatouage sera justifié par la suite, au regard de son impact sur le système de tatouage.

Le seuil MPEG

Parmi les modèles psychoacoustiques développés dans le cadre de MPEG, nous nous sommes plus particulièrement intéressés au modèle psychoacoustique n°1 de MPEG. Ce modèle se base

5. Le MPA n°1 de MPEG cherche à détecter précisément les composantes tonales et non tonales du signal audio avant d'appliquer les fonctions d'étalement avec les valeurs spécifiques de l'indice de masquage en fonction de la configuration (tonale ou non-tonale) détectée ; le MPA n°2 (utilisé dans le codeur Audio Advanced Coding (AAC)) effectue une analyse globale de la tonalité (via une mesure d'aplatissement/étalement spectrale) qui détermine un indice de masquage identique pour toutes les composantes fréquentielles de la trame de signal analysé.

6. Les paramètres des modèles sont réglés de sorte que le rapport masque à signal audio soit égal à −20 dB.



(a) (b)

Figure 5.4 – Allure du seuil de masquage obtenu avec trois modèles psychoacoustiques différents (le modèle MPEG n°1 [Nor93], le modèle de Gomes [Gom02] et celui de Garcia [Gar99]) pour une trame de 512 échantillons d'un signal de violon échantillonné à $44,1$ kHz : (a) allure temporelle du signal audio, (b) DSP du signal audio $S_x(f)$ et seuils de masquage associé $S_M(f)$.

sur une estimation de la DSP du signal audio, calculée sur une fenêtre du signal après avoir ajusté son niveau d'intensité à 96 dB SPL. Cette DSP est décomposée en sous-bandes correspondant plus ou moins aux bandes critiques. Dans chaque sous-bande, les composantes tonales et non-tonales significatives sont sélectionnées. Les fonctions d'étalement de chacune de ces composantes sont ensuite calculées pour être ajoutées les unes aux autres et définir le seuil de masquage global. Ce seuil est ensuite modifié pour prendre en compte le seuil d'audition absolu. Pour finir, il est redécomposé en sous-bandes, correspondant cette fois aux 32 sous-bandes d'un banc de filtres pseudo-QMF, pour ne conserver dans chaque sous-bande pseudo-QMF que sa valeur minimum, donnant à la courbe de masquage une allure discontinue caractéristique (dont un exemple est donné figure 5.4).

Ce seuil de masquage peut être appliqué au contexte du tatouage à condition de prendre deux précautions : d'une part supprimer la prise en compte du seuil d'audition absolu dans l'ajustement du seuil (surtout dans les hautes fréquences) et d'autre part adapter le rapport de puissance entre le seuil de masquage et le signal audio.
En effet, le seuil de masquage de MPEG est destiné au processus d'allocation binaire du codeur MPEG. Cette allocation est effectuée sous-bande par sous-bande en sélectionnant celle dont le rapport de puissance entre le masque et le bruit de quantification (à une étape donnée de l'algorithme) est le moins favorable [Nor93]. De fait, les sous-bandes hautes fréquences pour lesquelles la DSP du signal audio est très souvent inférieure au seuil d'audition absolu ne reçoivent la plupart du temps pas de bit lors de l'allocation. L'information audio dans ces sous-bandes n'est donc pas restituée après compression. Le problème du tatouage diffère ici du problème de compression puisqu'il s'agit d'ajouter une information inaudible. Placer l'information à tatouer, même (voire surtout) en choisissant sa DSP juste inférieure au seuil d'audition absolu, c'est prendre le risque de voir cette information perdue après compression. Pour éviter de concentrer la puissance de l'information à tatouer dans les hautes fréquences, il faut donc supprimer la prise en compte du seuil d'audition absolu dans l'ajustement du seuil de masquage.
De plus, le seuil de masquage est défini à un facteur de puissance près, et ce pour deux raisons.

D'une part, la puissance du signal audio avant évaluation du seuil de masquage est modifiée de sorte que la valeur maximale de sa DSP atteigne 96 dB SPL, simulant ainsi des conditions d'écoute spécifiques [7]. D'autre part, le processus d'allocation s'arrête non pas lorsque la limite d'inaudibilité du bruit de quantification est atteinte, mais lorsque la totalité des bits alloués pour le codage du signal sur une trame est atteinte. Le seuil de masquage influence le nombre de bits alloués à chaque sous-bande. Il confère au bruit de quantification une allure fréquentielle similaire à celle du seuil de masquage mais n'impose pas le rapport bruit à masque obtenu en fin de quantification. En conséquence, le rapport de puissance existant entre le seuil de masquage et le signal audio n'a pas d'influence sur l'algorithme de codage. Il n'a donc pas été optimisé dans le cadre des travaux MPEG. Par contre, dans un contexte de tatouage, ce rapport de puissance à la limite de l'inaudibilité est nécessaire. Un facteur de gain α paramétrant le rapport de puissance entre le seuil de masquage et la DSP du signal audio devra donc être ajouté au modèle psychoacoustique non optimisé si l'on souhaite utiliser ce modèle dans un contexte de tatouage.

Le seuil de Gomes

Le seuil de masquage proposé par Gomes est une version simplifiée du modèle n°1 de MPEG. La DSP du signal audio est décomposée cette fois en 4 sous-bandes. Dans chacune de ces sous-bandes, la DSP subit une compression de dynamique, une opération de convolution par une unique fonction d'étalement (sans distinction de tonalité des composantes) puis un lissage. Le seuil obtenu est ensuite translaté pour obtenir un Rapport de puissance entre le seuil de Masquage et le Signal audio (RMS) acceptable. Ce RMS influant sur les conditions d'inaudibilité, cette translation (dans le domaine fréquentielle) est paramétrable : elle prend d'un facteur de gain α dans le domaine temporel. Ce seuil offre finalement l'avantage d'une très faible complexité de calcul (par rapport notamment au modèle MPEG).

Le seuil de Garcia

Le modèle psychoacoustique de Garcia s'inspire du modèle psychoacoustique n°2 de MPEG. Garcia propose de calculer le seuil de masquage sur l'échelle des bandes critiques afin de respecter au mieux la physiologie de la perception auditive. La DSP du signal audio est transposée dans l'échelle des Bark par intégration de la puissance dans chaque sous-bande. Une fonction d'étalement unique (sans distinction de tonalité des composantes) est appliquée pour modéliser les effets de masquage de la membrane basilaire. La courbe obtenue est ensuite atténuée par un indice de masquage, variant avec la fréquence et dépend d'une mesure de tonalité du signal (basée sur le coefficient d'aplatissement du signal). Après atténuation, la courbe est modifiée pour prendre en compte le seuil d'audition absolu, donnant lieu au seuil de masquage.

5.1.2 Filtre de mise en forme spectrale

En utilisant un modèle psychoacoustique adapté, il est donc possible d'établir les conditions d'inaudibilité que doit respecter le signal modulé $v(n)$ pour être transparent s'il est ajouté au signal audio $x(n)$. Ces conditions prennent la forme d'un seuil de masquage homogène à une DSP qui traduit la limite fréquentielle d'inaudibilité à un facteur multiplicatif α près. Le signal $v(n)$ ajouté à $x(n)$ est transparent à condition que sa DSP soit inférieure au seuil de masquage

7. Comme l'oreille a un comportement fortement non linéaire, le seuil de masquage devrait être déterminé en fonction des conditions d'écoutes au récepteur !

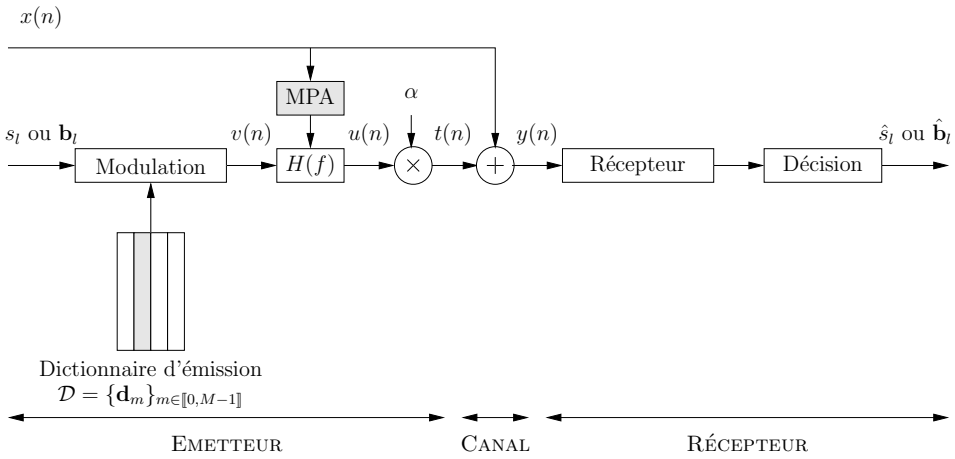


Figure 5.5 – Système de tatouage lorsque l'émetteur réalise la mise en forme perceptuelle du signal modulé en suivant un MPA.

quelle que soit la fréquence. Ce modèle psychoacoustique ne peut être appliqué que sur une fenêtre de durée N_H du signal audio, sur laquelle il peut être supposé stationnaire. Cette durée est de l'ordre de quelques millisecondes. L'analyse psychoacoustique doit donc être rééditée à chaque fois que les propriétés statistiques du signal audio sont modifiées.

Le choix d'un signal modulé $v(n)$ adapté (et implicitement le choix des vecteurs du dictionnaire d'émission) dépend de ce seuil de masquage. Deux conditions doivent être conciliées au regard de ce seuil. D'une part, la théorie des communications numériques [Pro01] établit que la probabilité d'erreur est minimum pour un RSB maximal. Ce RSB est celui obtenu en amont du corrélateur ; il dépendra donc des traitements effectués par le récepteur. Néanmoins, nous pouvons supposer que plus le rapport de puissance entre le signal modulé $v(n)$ et le signal audio $x(n)$ est grand, meilleur est le RSB en amont du corrélateur. La puissance du signal modulé $v(n)$ doit donc être choisie la plus élevée possible pour faciliter la détection correcte de l'information émise. D'autre part, cette puissance est limitée par le seuil de masquage puisque la DSP de $v(n)$, notée $S_v(f)$, ne doit pas excéder le seuil. La solution optimale de ce problème de maximisation sous contrainte consiste naturellement à choisir le signal modulé $v(n)$ de sorte que sa DSP coïncide avec le seuil de masquage. Or $v(n)$ est construit à partir d'un dictionnaire d'émission, fixé une fois pour toute, alors que le seuil de masquage est variable dans le temps $S_M(f)$. Il est donc préférable de réaliser la mise en forme spectrale de $v(n)$ par un filtre adaptatif, appelé ***filtre de mise en forme spectrale*** et noté $H(f)$. Ce filtre est construit de sorte que la DSP du signal résultant du filtrage de $v(n)$ par $H(f)$, appelé ***signal modulé mis en forme spectralement*** $u(n)$, coïncide avec le seuil de masquage. Cette relation de filtrage se traduit par :

$$S_u(f) = S_M(f) = |H(f)|^2 S_v(f), \tag{5.3}$$

où $S_u(f)$ est la densité spectrale de puissance de $u(n)$. Le ***signal de tatouage*** $t(n)$ est ensuite déduit en appliquant le facteur de gain α contrôlant le niveau d'inaudibilité offert par le seuil de masquage :

$$t(n) = \alpha u(n) = \alpha h(n) \star v(n), \tag{5.4}$$

où $h(n)$ est la réponse impulsionnelle du filtre de mise en forme $H(f)$ et $\star$ désigne le produit de convolution. La modulation prend alors la forme présentée figure 5.5.

Le filtre de mise en forme $H(f)$ peut être facilement évalué si $v(n)$ est choisi de sorte que sa DSP $S_v(f)$ soit constante quelle que soit la fréquence : en effet, si $S_v(f) = \sigma_v^2$, alors $H(f) = \dfrac{S_M(f)}{\sigma_v^2}$, autrement dit $H(f)$ est le filtre dont la réponse en fréquence approche au mieux le seuil de masquage à une constante multiplicative près. Cette constante multiplicative peut être prise en charge par le facteur multiplicatif α. On peut donc choisir $\sigma_v^2 = 1$.

La détermination des coefficients du filtre de mise en forme spectrale peut être le résultat d'une méthode de synthèse de filtre quelconque. Une méthode particulièrement facile à implémenter, et déjà utilisée par Gomes dans [Gom02], consiste à exploiter les techniques de modélisation paramétrique des signaux par des modèles autorégressifs [Hay96]. Cette procédure fournit un filtre, tout-pôle, toujours stable, causal et à déphasage minimal mais non linéaire, sous la forme :

$$H(z) = \frac{b_0}{1 + \sum_{i=1}^{P_H} a_i z^{-i}}, \tag{5.5}$$

où P_H est l'ordre du modèle autorégressif à paramétrer, les a_i sont les coefficients et b_0 le gain de $H(f)$.

Connaissant le seuil de masquage $S_M(f)$, nous pouvons évaluer la fonction d'autocorrélation r_M qui lui est associée à l'aide de la transformée de Fourier discrète inverse :

$$\forall n \in [\![0, N_H - 1]\!], r_M(n) = \frac{1}{N_H} \sum_{k=0}^{N_H-1} S'_M(k) \exp\left(j \frac{2\pi}{N_H} kn\right), \tag{5.6}$$

où $S'_M(k)$ est l'échantillonnée fréquentielle du seuil de masquage. L'objectif étant de faire coïncider la DSP de $u(n)$ et le seuil de masquage, cette fonction d'autocorrélation est également celle du signal $u(n)$. Elle peut donc être introduite à l'aide de l'équation de filtrage de $u(n)$. En effet, en posant $a_0 = 1$, on déduit :

$$u(n) = b_0 v(n) - \sum_{i=1}^{P_H} a_i u(n-i) \Rightarrow \forall l, \sum_{i=0}^{P_H} a_i r_M(l-i) = E\left[b_0 v(n) u(n-l)\right]. \tag{5.7}$$

Le terme $E[v(n)u(n-l)]$ est nul dès lors que $l > 0$, puisque les valeurs passées du signal $u(n-l)$ en sortie du filtre pour $l > 0$ ne dépendent pas de la valeur actuelle de l'entrée $v(n)$. Ainsi :

$$\forall l > 0, \sum_{i=0}^{P_H} a_i r_M(l-i) = 0 \Rightarrow \sum_{i=1}^{P_H} a_i r_M(l-i) = -r_M(l). \tag{5.8}$$

En développant cette relation sous la forme d'un système de P_H équations, nous obtenons les équations de Yule-Walker sous forme matricielle :

$$\Gamma_M \begin{bmatrix} a_1 \\ \vdots \\ a_{P_H} \end{bmatrix} = -\gamma_M, \tag{5.9}$$

où Γ_M désigne la matrice d'autocovariance de dimension $P_H \times P_H$ du signal $u(n)$ et γ_M est le vecteur d'autocovariance, soit :

$$\Gamma_M = \begin{bmatrix} r_M(0) & r_M(1) & \dots & r_M(P_H-1) \\ r_M(1) & r_M(0) & \dots & r_M(P_H-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_M(P_H-1) & r_M(P_H-2) & \dots & r_M(0) \end{bmatrix} \text{ et } \gamma_M = \begin{bmatrix} r_M(1) \\ r_M(2) \\ \vdots \\ r_M(P_H) \end{bmatrix}. \tag{5.10}$$

Ainsi, en représentant les coefficients a_i sous forme matricielle, nous déduisons :

$$\mathbf{a} = -\,{\Gamma_M}^{-1}\,\gamma_M. \tag{5.11}$$

Ce système d'équation peut être résolu de manière efficace à l'aide de l'algorithme de Durbin-Levinson [RJ93], en exploitant la structure Toeplitz de la matrice Γ_M. Le gain du filtre b_0 est déduit en utilisant la relation (5.7) pour $l = 0$:

$$b_0 = \sqrt{\sum_{i=0}^{P_H} a_i r_M(i)}. \tag{5.12}$$

5.1.3 Contrainte d'inaudibilité

Ce formalisme traduit à lui seul les conditions d'inaudibilité. Si $H(f)$ et α sont correctement paramétrés, choisir n'importe quel signal modulé de DSP constante et de puissance σ_v^2 égale à 1 garantit l'inaudibilité du signal de tatouage $t(n)$ en présence de $x(n)$. Cette condition peut être étendue au signal modulé de puissance inférieure à 1. En effet, dans ce cas, l'allure fréquentielle du signal de tatouage $t(n)$ est toujours celle du seuil de masquage (au facteur de gain près), mais la tangence entre le seuil et la DSP de $t(n)$ n'est pas atteinte. Le signal $t(n)$ est moins puissant que le seuil et est là encore inaudible. L'inaudibilité du tatouage est donc satisfaite si le signal modulé $v(n)$ est choisi blanc et de puissance :

$$\sigma_v^2 \leq 1. \tag{5.13}$$

Cette condition sera par la suite appelée ***contrainte d'inaudibilité***.

5.1.4 Modulation par étalement de spectre

Choisir $v(n)$ de DSP constante revient à opter pour un signal modulé blanc étalé en fréquence. Ce signal étant construit à partir du dictionnaire, les vecteurs du dictionnaire d'émission sont donc choisis :

- blancs, étalés dans la bande de fréquence $[0; F_c]$, où F_c est la largeur de bande à définir. Cette largeur de bande peut pour l'instant être choisie égale à $\dfrac{F_e}{2}$ si l'on souhaite utiliser toute la gamme de fréquences disponibles pour des signaux échantillonnés à F_e,
- de même puissance unité :

$$\forall m \in [\![0, M-1]\!], \frac{\mathbf{d}_m^t \mathbf{d}_m}{N_s} = 1. \tag{5.14}$$

- et orthogonalisé (ou rendu bi-orthogonaux) pour les besoins du récepteur par la procédure d'orthonormalisation de Gram-Schmidt.

Ainsi, quelle que soit l'information émise, le signal modulé $\mathbf{v}$ construit sur chaque temps symbole est de puissance unité :

$$\sigma_v^2 = \frac{\mathbf{v}^t \mathbf{v}}{N_s} = 1, \tag{5.15}$$

et satisfait à la contrainte d'inaudibilité.

Le signal modulé $v(n)$ est finalement le résultat d'une ***modulation par étalement de spectre***. Ce type de modulation offre un avantage non négligeable pour la chaîne de tatouage. Utiliser des vecteurs étalés spectralement permet d'introduire une redondance de l'information, connue pour protéger efficacement le système de transmission s'il est soumis à des interférences sévères [Pro01] (ce qui est notre cas). Les modulations par étalement de spectre sont en effet souvent utilisées pour les transmissions satellitaires par exemple dans lesquelles les rapports signaux à bruit sont défavorables à la transmission. Outre la redondance, leur aspect pseudo-aléatoire rend difficile la démodulation/détection de l'information tatouée à un récepteur non autorisé [Pro01]. Cette propriété s'avère particulièrement intéressante pour les systèmes de tatouage dédiés à la protection de la propriété intellectuelle et justifie l'utilisation des modulations par étalement dans de nombreux systèmes de la littérature.

5.2 Réception

Cette section s'intéresse à la conception du récepteur de la chaîne de tatouage. Deux hypothèses sont effectuées : d'une part, aucune perturbation supplémentaire n'est introduite par le canal de transmission de sorte que le signal reçu en sortie de l'émetteur s'exprime sous la forme :

$$y(n) = x(n) + t(n) = x(n) + \alpha h(n) \star v(n). \tag{5.16}$$

D'autre part, l'IES est négligée ; la détection pourra être effectuée en observant le signal reçu sur chaque temps symbole.

Le récepteur optimal défini par la théorie des communications numériques ne peut être appliqué directement à la chaîne de tatouage. En effet, il suppose un bruit $x(n)$ blanc Gaussien, très éloigné des caractéristiques du signal audio. Ce signal est supposé modélisable par un processus AR d'ordre P_0 et localement stationnaire sur des fenêtres de durée N_G (inférieures à 20 ms). De plus, un filtre de mise en forme perceptuelle a été introduit à l'émetteur pour les besoins de la contrainte d'inaudibilité. Ce filtre modifie les caractéristiques du signal de tatouage $t(n)$ et ceci doit être pris en compte par le récepteur. Un récepteur adapté à la statistique du signal audio et à l'émetteur de la chaîne de tatouage doit donc être conçu.

5.2.1 Récepteur sous-optimal pour un bruit auto-régressif

5.2.1.1 Prise en compte du filtre de mise en forme au récepteur

Supposons le signal reçu $y(n) = x(n)+\alpha h(n)\star v(n)$ observé sur une fenêtre de N_G échantillons. Notons $\mathbf{y}$, $\mathbf{v}$ et $\mathbf{x}$ l'observation des signaux $y(n)$, $v(n)$ et $x(n)$ sur cette fenêtre. Par hypothèse, la composante audio $\mathbf{x}$ peut être modélisée par un processus aléatoire $X(n)$ stationnaire, ergodique, Gaussien, centré et AR d'ordre P_0. Nous noterons Γ_X la matrice d'autocovariance de dimension $N_G \times N_G$ du processus $X(n)$. Étant donnée l'introduction à l'émetteur du filtre de mise en forme perceptuelle $H(f)$, l'observation du signal reçu $\mathbf{y}$ est la somme de la composante aléatoire $\mathbf{x}$ et d'une composante déterministe $H\mathbf{v}$. $H\mathbf{v}$ est une représentation vectorielle de l'opération de

filtrage de $v(n)$ par $H(f)$ avec :

$$H = \begin{bmatrix} h(0) & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & h(0) & \ddots & & & \vdots \\ h(P_h) & \vdots & \ddots & \ddots & & \vdots \\ 0 & h(P_h) & & h(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & h(P_h) & \cdots & h(0) \end{bmatrix}, \tag{5.17}$$

à supposer la réponse impulsionnelle $h(n)$ de $H(f)$ tronquée à l'ordre P_h.

La loi de probabilité de $\mathbf{Y}$ sachant le vecteur $\mathbf{v}$ émis a maintenant pour densité de probabilité :

$$p_{\mathbf{Y}|\mathbf{v}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{N_G}{2}}\sqrt{\det(\Gamma_X)}} \exp\left(-\frac{1}{2}(\mathbf{y}-\alpha H\mathbf{v})^t (\Gamma_X)^{-1}(\mathbf{y}-\alpha H\mathbf{v})\right). \tag{5.18}$$

5.2.1.2 Mise en évidence du filtre blanchissant

Supposer $X(n)$ autorégressif d'ordre P_0 revient à écrire :

$$\tilde{X}(n) = X(n) + \sum_{i=1}^{P_0} g_i X(n-i), \tag{5.19}$$

où $\tilde{X}(n)$ est un processus aléatoire blanc Gaussien et de puissance $\tilde{\sigma}_x^2$. Filtrer le signal audio $x(n)$ par le filtre[8] $G(z) = 1 + \sum_{i=1}^{P_0} g_i z^{-i}$ permet donc de se ramener à un bruit $\tilde{x}(n)$ blanc et Gaussien. Ce filtre est donc introduit dans la chaîne de tatouage en guise de pré-traitement du signal reçu $y(n)$. En déplaçant l'opération de filtrage en amont de l'additionneur, comme le montre la figure 5.6, le système de tatouage est alors ramené à une configuration de chaîne de communication avec BBAG : le signal porteur de l'information tatouée est $w(n)$ résultat du filtrage du signal de tatouage $t(n)$ par $G(f)$ et le bruit de canal est $\tilde{x}(n)$, c'est à dire un bruit BBAG.

Il convient alors d'exprimer la loi de probabilité de l'observation $\mathbf{y}$ en fonction du bruit de canal $\tilde{\mathbf{x}}$. Dans cet objectif, l'expression de la matrice d'autocovariance Γ_X de $\mathbf{X}$ peut aisément être établie en fonction du filtre blanchissant et de la matrice d'autocovariance $\Gamma_{\tilde{X}}$ du vecteur aléatoire $\tilde{\mathbf{X}}$. Partant de conditions initiales nulles, l'opération de filtrage liant $\tilde{x}(n)$ et $x(n)$ peut être représentée sous forme matricielle :

$$\tilde{\mathbf{x}} = G\mathbf{x} \text{ avec } G = \begin{bmatrix} 1 & 0 & \cdots & & & \cdots & 0 \\ g_1 & 1 & \ddots & & & & \vdots \\ \vdots & g_1 & \ddots & & & & \\ g_{P_0} & \vdots & \ddots & 1 & & & \\ 0 & g_{P_0} & & g_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & g_{P_0} & \cdots & g_1 & 1 \end{bmatrix}. \tag{5.20}$$

8. Par la suite, nous représenterons le filtre $G(z)$ par $G(f)$ avec $z = \exp(j2\pi f)$ pour simplifier les notations.

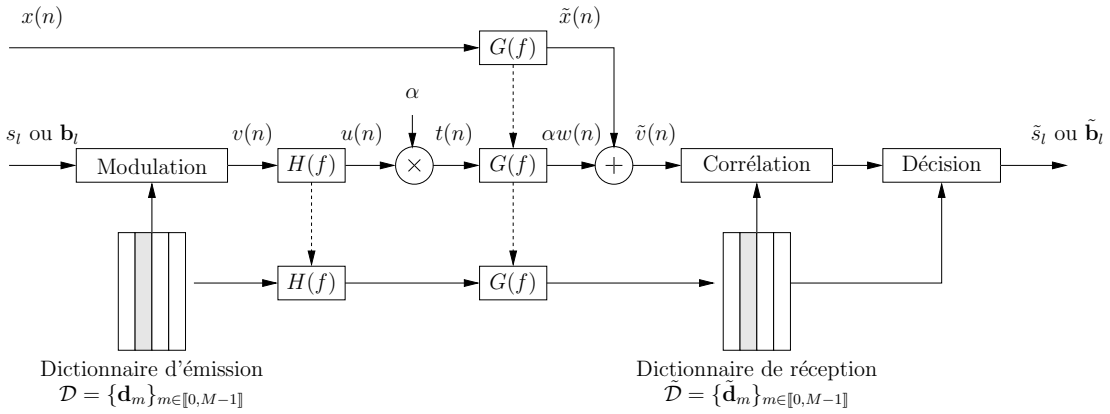


Figure 5.6 – Introduction du récepteur sous-optimal dans le système de tatouage.

Il suffit alors d'utiliser l'opération de filtrage inverse, qui s'écrit : $\mathbf{X} = G^{-1}\tilde{\mathbf{X}}$. La matrice d'autocovariance d'un vecteur aléatoire $\mathbf{X}$ étant par définition $\Gamma_X = E[\mathbf{X}\mathbf{X}^t]$, nous en déduisons :

$$\Gamma_X = G^{-1}\Gamma_{\tilde{X}}\left(G^{-1}\right)^t. \tag{5.21}$$

Or les N_G composantes de $\tilde{\mathbf{X}}$ sont indépendantes (puisque $\tilde{X}(n)$ est blanc et Gaussien) et suivent la même loi $\mathcal{N}(0, \tilde{\sigma}_x^2)$, ainsi $\Gamma_{\tilde{X}}$ est une matrice diagonale pour laquelle tous les termes diagonaux sont $\tilde{\sigma}_x^2$. Nous en déduisons l'expression de l'inverse de la matrice d'autocovariance de $X(n)$ en fonction de la représentation matricielle du filtrage blanchissant :

$$(\Gamma_X)^{-1} = \frac{G^t G}{\tilde{\sigma}_x^2}. \tag{5.22}$$

Au regard de l'expression matricielle précédente, la loi de probabilité de $\mathbf{Y}$ sachant le vecteur $\mathbf{v}$ émis devient :

$$p_{\mathbf{Y}|\mathbf{v}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{N_G}{2}}\sqrt{\det(\Gamma_X)}}\exp\left(-\frac{1}{2}\frac{||G\mathbf{y} - \alpha GH\mathbf{v}||^2}{\tilde{\sigma}_x^2}\right). \tag{5.23}$$

5.2.1.3 Métriques de détection

Introduisons $\mathbf{w} = GH\mathbf{v}$ le vecteur résultant du filtrage de $v(n)$ par l'opération de mise en forme spectrale associée à la matrice H et celle de blanchiment associée à la matrice G. Le vecteur $\mathbf{v}$ ne pouvant prendre que M^L états finis [9], $\mathbf{w}$ ne peut prendre que les 2^{KL} formes résultant du filtrage des 2^{KL} formes possibles de $\mathbf{v}$ par le filtre de mise en forme perceptuelle $H(f)$ et le filtre blanchissant $G(f)$. Les états de $\mathbf{w}$, de la même façon que les états de $\mathbf{v}$, peuvent être modélisés par une variable aléatoire $W = \{0, \ldots, 2^{KL} - 1\}$. Étant donné l'opération de filtrage liant $\mathbf{v}$ et $\mathbf{w}$, $\mathbf{v}$ est dans l'état $V = m$ si et seulement si $\mathbf{w}$ est dans l'état $W = m$ [10].

Le critère du maximum de vraisemblance, consistant à rechercher l'état de $\mathbf{v}$ qui maximise les densités de probabilité $p_{\mathbf{Y}|\mathbf{v}}(\mathbf{y})$, est donc équivalent à trouver l'état $\hat{m}$ de $\mathbf{w}$ qui maximise les

9. où L désigne ici le nombre de symboles ou K-uplet binaire transmis pendant la durée N_G.
10. Nous supposerons que les filtres ne modifient pas la dimension de l'espace signal.

densités de probabilités de $\mathbf{Y}$ connaissant $\mathbf{w}$:

$$\hat{m} = \arg \max_{m \in [\![0, 2^{KL}-1]\!]} p_{\mathbf{Y}|\mathbf{w}}(\mathbf{y}),$$
$$\text{avec } p_{\mathbf{Y}|\mathbf{w}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{N_G}{2}} \sqrt{\det(\Gamma_X)}} \exp\left(-\frac{1}{2} \frac{||G\mathbf{y} - \alpha\mathbf{w}||^2}{\tilde{\sigma}_x^2}\right) \tag{5.24}$$

Nous retrouvons l'expression de la loi de probabilité du détecteur optimal dans le cas BBAG lorsque le signal observé $\tilde{\mathbf{v}}$ est le signal $G\mathbf{y}$, résultant du filtrage du signal reçu $y(n)$ par le filtre blanchissant $G(f)$, et le signal porteur de l'information est $w(n)$. Le récepteur optimal consiste donc à évaluer les métriques de décision :

$$C'(\tilde{\mathbf{v}}, \mathbf{w}) = \tilde{\mathbf{v}}^t \mathbf{w} - \frac{1}{2}\alpha ||\mathbf{w}||^2, \tag{5.25}$$

pour tous les états de $\mathbf{w}$ possibles et à sélectionner l'état qui maximise ces métriques.

5.2.1.4 Calcul du filtre blanchissant

Les coefficients g_i du filtre blanchissant sont obtenus en appliquant l'algorithme de Durbin-Levinson proposé par la théorie de la Prédiction Linéaire (PL) [Hay96]. Il consiste à rechercher les P_G coefficients du filtre $G(f)$, notés sous forme vectorielle $\mathbf{g} = [g_1, ..., g_{P_G}]^t$, qui minimise la puissance de l'erreur de prédiction $\tilde{x}(n)$:

$$\tilde{\sigma}_x^2 = E[\tilde{x}(n)^2], \text{ avec } \tilde{x}(n) = x(n) + \sum_{i=1}^{P_G} g_i x(n-i). \tag{5.26}$$

L'ordre P_G du filtre $G(f)$ est un paramètre du système qui sera établi par la suite.

En annulant la dérivée de la puissance de l'erreur de prédiction par rapport aux coefficients du filtre $G(f)$ nous retrouvons les équations de Wiener-Hopf :

$$\Gamma_x \begin{bmatrix} g_1 \\ \vdots \\ g_{P_G} \end{bmatrix} = -\gamma_x, \tag{5.27}$$

où Γ_x désigne la matrice d'autocovariance et γ_x le vecteur d'autocovariance de $x(n)$ d'ordre P_G, soit :

$$\Gamma_x = \begin{bmatrix} \gamma_x(0) & \gamma_x(1) & \dots & \gamma_x(P_G-1) \\ \gamma_x(1) & \gamma_x(0) & \dots & \gamma_x(P_G-2) \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_x(P_G-1) & \dots & \gamma_x(1) & \gamma_x(0) \end{bmatrix}, \gamma_x = \begin{bmatrix} \gamma_x(1) \\ \gamma_x(2) \\ \vdots \\ \gamma_x(P_G) \end{bmatrix}, \tag{5.28}$$

avec $\gamma_x(i) = E[x(n)x(n-i)]$ la fonction d'autocovariance de $x(n)$.

Les coefficients du filtre, sous forme vectorielle $\mathbf{g}$ se déduisent naturellement en inversant la matrice Γ_x par :

$$\mathbf{g} = -\,{\Gamma_x}^{-1}\,\gamma_x. \tag{5.29}$$

Ces coefficients sont calculés sur une fenêtre d'analyse de durée N_G où le signal audio $x(n)$ peut être supposé localement stationnaire. Ils doivent donc être réévalués tous les N_G échantillons.

5.2.1.5 Introduction du dictionnaire de réception

Le signal reçu $\tilde{v}(n)$, après filtrage de $y(n)$ par le filtre blanchissant $G(f)$, est maintenant observé sur chaque temps symbole de durée N_s. Les états du vecteur $\mathbf{w}$ à détecter se réduisent au nombre de 2^K. Il s'agit en effet des 2^K formes du signal modulé construites à partir des M vecteurs du dictionnaire d'émission filtrés par le filtre blanchissant $G(f)$. Ces formes dépendent de la modulation choisie :

- pour un message numérique sous la forme d'une suite de symboles,

$$\mathbf{v} = \mathbf{d}_{s_l} \Rightarrow \mathbf{w} = GH\mathbf{d}_{s_l} \tag{5.30}$$

- pour un message numérique sous forme d'une suite de vecteurs, la linéarité du filtrage permet d'établir :

$$\mathbf{v} = \sum_{k=0}^{K}(2\mathbf{b}_l[k]-1)\mathbf{d}_k \Rightarrow \mathbf{w} = \sum_{k=0}^{K}(2\mathbf{b}_l[k]-1)GH\mathbf{d}_k. \tag{5.31}$$

Ces 2^K formes peuvent toutes être déduites d'un ***dictionnaire de réception***, de dimension $N_s \times M$, contenant l'ensemble des formes d'onde du dictionnaire d'émission filtrées par le filtre blanchissant $G(f)$. Ce dictionnaire de réception doit être réévalué sur chaque temps symbole N_s (bien évidemment différent des durées N_H et N_G sur lesquelles sont évaluées le filtre de mise en forme perceptuelle $H(f)$ et le filtre blanchissant $G(f)$). Il sera noté :

$$\tilde{\mathcal{D}} = \left\{\tilde{\mathbf{d}}_m\right\}_{m\in[\![0,M-1]\!]}, \text{ avec } \forall m, \tilde{\mathbf{d}}_m = GH\mathbf{d}_m. \tag{5.32}$$

Le vecteur $\mathbf{w}$ peut finalement être obtenu en modulant l'information à émettre à l'aide du dictionnaire de réception, de la même façon que le signal modulé $\mathbf{v}$ a été construit à l'aide du dictionnaire d'émission. La chaîne devient équivalente à un canal BBAG pour lequel le signal modulé est le signal $\mathbf{w}$ construit à partir du dictionnaire de réception et le bruit est le signal audio blanchi par le filtre $G(f)$, noté $\mathbf{r}$.

Les expressions spécifiques du récepteur pour chaque type de modulation sont alors déduites en appliquant celles établies dans le cas BBAG au dictionnaire de réception $\tilde{\mathcal{D}}$ et au signal reçu $\tilde{\mathbf{v}}$:

- pour une modulation M-aire, la règle de décision consiste cette fois à rechercher l'indice du dictionnaire de réception qui maximise les métriques de détection $C'(\tilde{\mathbf{v}}, \mathbf{w}) = \tilde{\mathbf{v}}^t\mathbf{w} - \frac{1}{2}\alpha||\mathbf{w}||^2$, avec $\mathbf{w} = \tilde{\mathbf{d}}_m$. Le symbole décidé $\tilde{s}_l$ est donc :

$$\tilde{s}_l = \arg\max_{m\in[\![0,M-1]\!]} C'(\tilde{\mathbf{v}}, \tilde{\mathbf{d}}_m), \text{ avec } C'(\tilde{\mathbf{v}}, \tilde{\mathbf{d}}_m) = \tilde{\mathbf{v}}^t\tilde{\mathbf{d}}_m - \frac{1}{2}\alpha||\tilde{\mathbf{d}}_m||^2. \tag{5.33}$$

  Cette expression pourrait conduire à un détecteur par corrélation à la condition que les vecteurs du dictionnaire de réception soient de même norme. Le respect de cette condition sera étudié par la suite.
  Dans tous les cas, nous pourrons simplifier le détecteur en normalisant les vecteurs de réception. Nous parlerons alors de détecteur par corrélation normalisée, comme proposé dans [MB99]. Le symbole décidé sera alors :

$$\tilde{s}_l = \arg\max_{m\in[\![0,M-1]\!]} C(\tilde{\mathbf{v}}, \tilde{\mathbf{d}}_m), \text{ avec } C(\tilde{\mathbf{v}}, \tilde{\mathbf{d}}_m) = \tilde{\mathbf{v}}^t\frac{\tilde{\mathbf{d}}_m}{||\tilde{\mathbf{d}}_m||}. \tag{5.34}$$

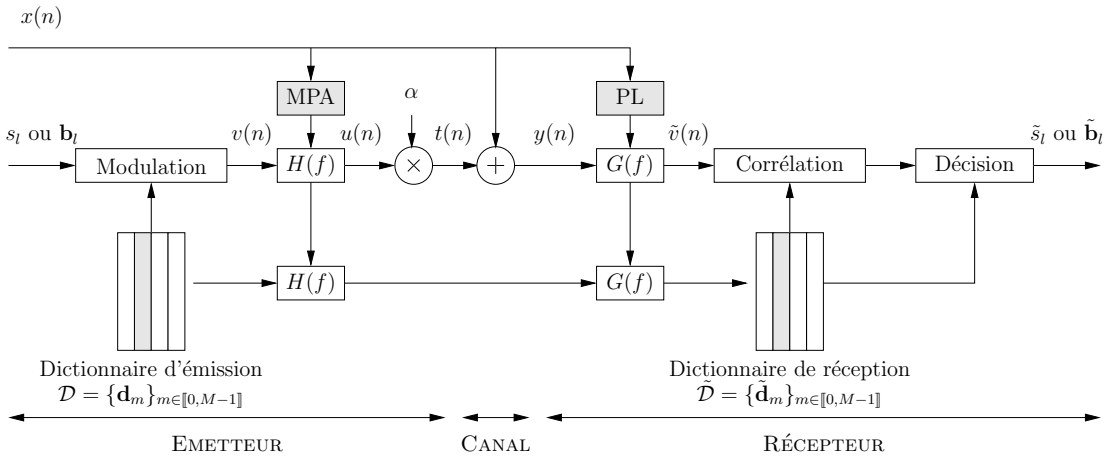


Figure 5.7 – Schéma du système de tatouage lorsque le récepteur sous-optimal est introduit.

- dans le cas d'une modulation CDMA, le signal $\mathbf{w}$ est une combinaison linéaire des vecteurs du dictionnaire de réception $\hat{\mathcal{D}}$ dont les coefficients sont ± 1. Le détecteur sous-optimal consiste donc à projeter le signal reçu sur les vecteurs du dictionnaire de réception puis à inverser les intercorrélations des vecteurs du dictionnaire de réception en exploitant sa matrice d'autocorrélation $\Gamma_{\tilde{\mathcal{D}}}$. Le K-uplet binaire reçu $\hat{\mathbf{b}}_l$ est donc donné par :

$$\tilde{\mathbf{b}}_l = \frac{\tilde{\mathbf{a}}_l + 1}{2}, \text{ avec } \tilde{\mathbf{a}}_l = \Gamma_{\tilde{\mathcal{D}}}^{-1} \mathcal{D}^t \tilde{\mathbf{v}}. \tag{5.35}$$

Le schéma de la chaîne de tatouage à ce stade du développement est présenté figure 5.7.

5.2.2 Vers un schéma de réception aveugle

Le filtre $G(f)$ est par construction le filtre permettant de blanchir le signal audio original $x(n)$. Ce signal n'est malheureusement pas disponible lors de la procédure de détection que nous avons souhaité aveugle. Ce filtre doit donc être estimé à partir du signal audio tatoué reçu $y(n)$ (puis par la suite $\hat{y}(n)$ lorsque les perturbations introduites par le canal seront considérées). De la même façon, l'évaluation du dictionnaire de réception met en œuvre le filtre de mise en forme $H(f)$, calculé après analyse psychoacoustique du signal audio $x(n)$. Là encore, le signal audio n'étant pas connu du récepteur, ce filtre doit être estimé à partir du signal audio tatoué reçu $y(n)$ (ou $\hat{y}(n)$ en cas de perturbation canal). Cette estimation est notée $\hat{H}(f)$.

La prise en compte du signal audio tatoué éventuellement dégradé pour le calcul des filtres $\hat{G}(f)$ et $\hat{H}(f)$ donne lieu au schéma de tatouage modifié présentée figure 5.8. Le signal reçu analysé par le corrélateur est désormais le signal $\hat{v}(n)$ avec le dictionnaire de réception effectif :

$$\hat{\mathcal{D}} = \left\{\hat{\mathbf{d}}_m\right\}_{m\in[\![0,M-1]\!]}, \text{ avec } \forall m, \hat{\mathbf{d}}_m = \hat{G}\hat{H}\mathbf{d}_m. \tag{5.36}$$

La décision prend donc la forme suivante :

- pour un message numérique sous la forme d'une suite de symboles, le symbole détecté est :

$$\hat{s}_l = \arg \max_{m\in[\![0,M-1]\!]} C'(\hat{\mathbf{v}}, \hat{\mathbf{d}}_m), \text{ avec } C'(\hat{\mathbf{v}}, \hat{\mathbf{d}}_m) = \hat{\mathbf{v}}^t \hat{\mathbf{d}}_m - \frac{1}{2}\alpha ||\hat{\mathbf{d}}_m||^2. \tag{5.37}$$

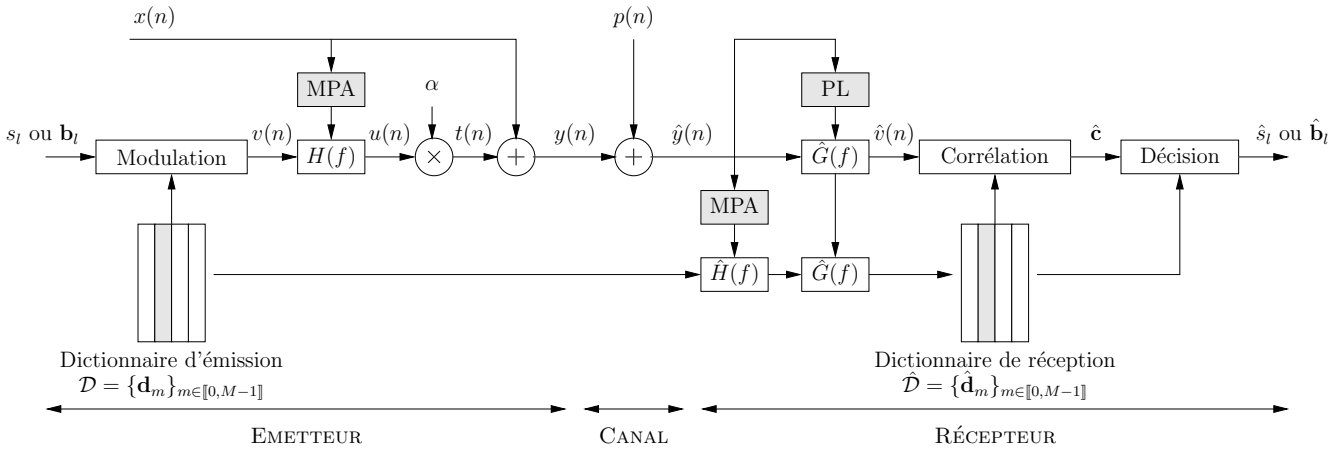


Figure 5.8 – Système de tatouage audio pour une détection aveugle de l'information de tatouage.

ou dans le cas où les vecteurs du dictionnaire sont de même norme,

$$\hat{s}_l = \arg \max_{m\in[\![0,M-1]\!]} C(\hat{\mathbf{v}}, \hat{\mathbf{d}}_m), \text{ avec } C(\hat{\mathbf{v}}, \hat{\mathbf{d}}_m) = \hat{\mathbf{v}}^t \hat{\mathbf{d}}_m. \tag{5.38}$$

- dans le cas d'une modulation CDMA, le K-uplet binaire détecté s'écrit sous forme vectorielle :

$$\hat{\mathbf{b}}_l = \frac{\hat{\mathbf{a}}_l + 1}{2}, \text{ avec } \hat{\mathbf{a}}_l = \Gamma_{\hat{\mathcal{D}}}^{-1} \hat{\mathcal{D}} \hat{\mathbf{v}}, \tag{5.39}$$

  où $\Gamma_{\hat{\mathcal{D}}}$ est la matrice de corrélation du dictionnaire de réception.

L'erreur introduite par l'estimation de $G(f)$ et de $H(f)$ à partir de $\hat{y}(n)$ laisse présager une dégradation des performances du système de tatouage si cette erreur est trop importante. Il conviendra donc d'évaluer l'impact de cette erreur d'estimation sur la probabilité d'erreur de la chaîne de tatouage. Cette étude sera effectuée par la suite.

Conclusion

Ce chapitre nous a permis d'établir le schéma du système de tatouage présenté figure 5.8. Ce système exploite les outils de la théorie des communications numériques en les adaptant aux spécificités et aux contraintes de l'application de tatouage. L'information à tatouer est modulée en exploitant une modulation par étalement de spectre de type M-aire ou CDMA. Elle est suivie d'une mise en forme spectrale qui confère au signal de tatouage $t(n)$ l'allure fréquentielle d'un seuil de masquage obtenu après une analyse psychoacoustique du signal audio ; ce seuil traduit la limite fréquentielle à satisfaire par un signal $t(n)$ pour être ajouté au signal audio $x(n)$ sans être audible. La réception se base sur un filtre blanchissant. Ce filtre permet de ramener la chaîne de tatouage à une configuration BBAG. Le détecteur est de nouveau un démodulateur par corrélation exploitant cette fois un dictionnaire de réception, résultat du filtrage du dictionnaire d'émission. Cette chaîne de tatouage a été établie sous certaines hypothèses (IES négligeable, orthonormalité du dictionnaire de réception, etc...) et à de nombreux paramètres près (choix du seuil de masquage, facteur d'amplitude α, bande passante du système, etc...) qu'il convient maintenant de justifier et de choisir.

Chapitre 6

Paramétrisation du système de tatouage

Le chapitre précédent nous a permis d'établir les éléments sur lesquels repose le système de tatouage : l'émetteur constitué d'une modulation par étalement de spectre suivie d'une mise en forme spectrale dépendante d'un seuil de masquage et d'un facteur de gain et le récepteur basé sur un filtre blanchissant et un démodulateur exploitant un calcul de corrélation. Ce système tolère pour l'instant un grand nombre de degrés de liberté, dont la liste exhaustive [1] est précisée table 6.1 : quel modèle psychoacoustique est le plus efficace, comment mettre en œuvre effectivement les filtres de la chaîne, quelle est la largeur de bande F_c adéquate des vecteurs du dictionnaire d'émission, etc... L'objectif de ce chapitre est de répondre à ces questions.

La paramétrisation du système est intiment liée aux performances du système : le choix de l'ordre des filtres influence la fiabilité de transmission, le choix du modèle psychoacoustique et du facteur d'amplitude α conditionne l'inaudibilité du tatouage ; il interfère également dans les conditions de transmission du signal de tatouage (rapport signal à bruit à l'entrée du corrélateur) qui détermine la fiabilité de transmission, la bande passante des vecteurs du dictionnaire d'émission conditionne la robustesse du système. Différents critères de performances seront donc proposés et mis en œuvre pour déterminer les paramètres les plus adaptés au système de tatouage.

6.1 Filtres de la chaîne de tatouage

6.1.1 Mise en oeuvre des filtrages

Deux filtres interviennent dans la chaîne de tatouage : le filtre de mise en forme perceptuelle $H(f)$ et le filtre blanchissant $G(f)$. Ces deux filtres sont évalués sur des fenêtres d'analyse du signal audio de durées respectives N_H et N_G où le signal audio peut être supposé localement stationnaire. Cette stationnarité locale est couramment établie pour des trames de l'ordre de 20 ms. Nous choisirons donc $N_H = N_G = 512$ échantillons. Ce choix simplifie le fenêtrage des signaux traités : en effet, la construction de la chaîne de tatouage établit trois types de

1. Le temps symbole N_s et le paramètre d'encodage K ne seront pas considérés dans cette paramétrisation, puisqu'ils induisent le débit de transmission $R = \frac{KF_e}{N_s}$ que nous ferons varier pour établir les performances du système.

Paramètres	Descriptions
$\mathbf{N_H}$	Durée des fenêtres d'analyse pour le calcul du MPA
$\mathbf{N_G}$	Durée des fenêtres d'analyse pour le calcul des filtres blanchissants
$\mathbf{P_H}$	Ordre du filtre $H(f)$
$\mathbf{P_G}$	Ordre du filtre $G(f)$
$\mathbf{F_c}$	Limite d'étalement pour les vecteurs du dictionnaire
α	Facteur d'amplitude
$\mathbf{S_M(f)}$	Seuil de masquage

Table 6.1 – Listes des paramètres de la chaîne de tatouage.

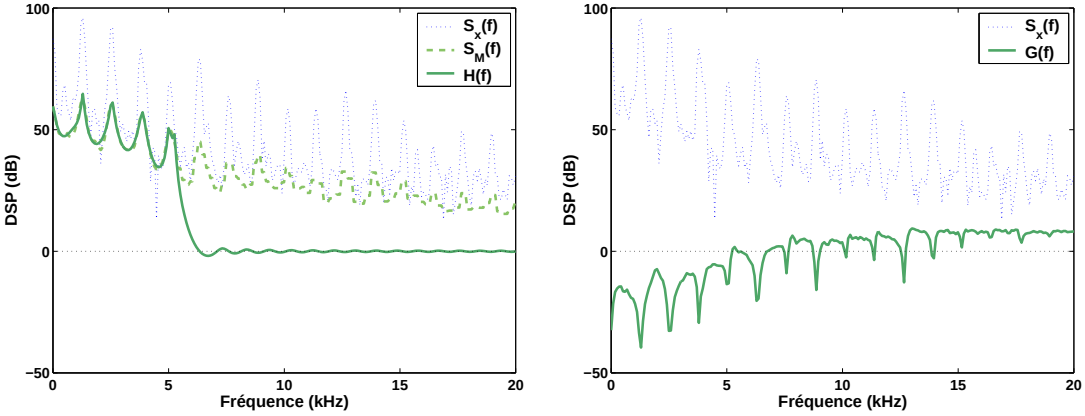
fenêtres d'analyse ; en plus de celles liées au calcul de $H(f)$ et $G(f)$, un fenêtrage découpant le signal en temps symbole N_s est effectué pour les besoins de la modulation et du détecteur par corrélation. Dans les simulations à venir, nous imposerons parfois un débit de transmission et un dictionnaire d'émission (ici égal à 71 bit/s pour un dictionnaire de 2 vecteurs) de sorte à avoir $N_s = N_H = N_G$, ce qui facilite le fenêtrage.

L'ordre de ces filtres doit également être établi. Gomes [Gom02] dans sa thèse s'est intéressé à la conception du filtre de mise en forme $H(f)$. Il montre que l'ordre du filtre a peu d'influence sur les performances du système (notamment en terme de TEB). Nous nous sommes donc confortés à ses conclusions et nous avons choisi pour ordre $P_H = 50$. La figure 6.1(a) présente à titre indicatif la réponse en fréquence du filtre $H(f)$ obtenu pour un seuil de masquage $S_M(f)$ (déjà présenté figure 5.4) dans le cas où ce seuil est calculé à l'aide du modèle psychoacoustique de Gomes pour une trame de signal de violon.

L'ordre du filtre doit être choisi de sorte à coïncider avec l'ordre du processus AR P_0 qui modélise le signal audio original $x(n)$. Néanmoins, P_0 varie d'un signal audio à un autre voire d'une fenêtre d'analyse à une autre. Il faudrait envisager d'adapter l'ordre du filtre P_G aux caractéristiques du signal. Ce procédé étant coûteux, nous avons préféré choisir une valeur de P_G fixe. La valeur optimale de P_G est couramment choisie pour obtenir un bon compromis entre la performance du filtre (c'est à dire sa capacité à minimiser l'erreur quadratique moyenne de prédiction) et le temps de calcul (lié au nombre de coefficients). Son choix fait alors appel à des critères de sélection de l'ordre du modèle, tel que le critère Minimum Description Length (MDL). Dans le contexte du tatouage, la performance du filtre relève de la minimisation du TEB de la chaîne de tatouage. Nous avons donc mesuré le TEB moyen de cette chaîne pour différentes valeurs de l'ordre du filtre blanchissant lorsque les signaux sont tatoués par une séquence de 5000 bits chacun à un débit de 71 bit/s. Les résultats obtenus sont présentés figure 6.2. La probabilité d'erreur atteint un minimum lorsque $P_G = 200$. Cette valeur de l'ordre sera utilisée par la suite. On notera néanmoins que le TEB fluctue relativement peu lorsque P_G varie entre 50 et 200 au regard de la précision des mesures ; si le système de tatouage est soumis à des contraintes de temps de calcul fortes, choisir un ordre de filtre de 50 pourra alors s'avérer être un bon compromis entre TEB et coût de mis en œuvre. L'allure de la réponse en fréquence du filtre blanchissant obtenu est présentée figure 6.1(b) pour une trame de signal de violon.

6.1.2 Influence de l'erreur d'estimation des filtres sur les performances

L'estimation de $G(f)$ et de $H(f)$ à partir de $\hat{y}(n)$ au lieu du signal audio original $x(n)$ laisse présager une modification des performances du système de tatouage. Nous nous sommes



(a) (b)

Figure 6.1 – Allure des réponses en fréquence du filtre de mise en forme $H(f)$ calculé d'après le MPA de Gomes (a) et du filtre blanchissant $G(f)$ (b) calculé sur une trame de signal de violon.

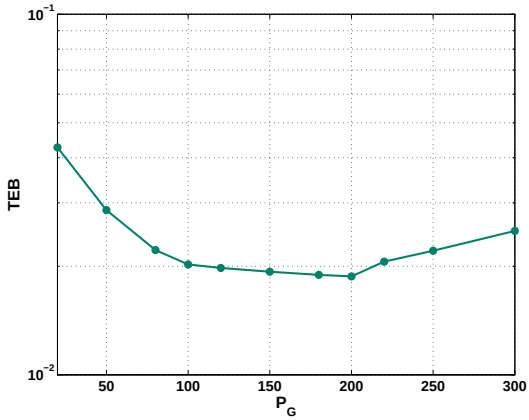


Figure 6.2 – Variation du TEB moyen en fonction de l'ordre P_G du filtre blanchissant. Valeurs moyennes obtenues sur l'ensemble des 20 signaux tests tatoués par 5000 bits chacun au débit de 71 bit/s pour un dictionnaire constitué de deux vecteurs antipodaux.

proposés d'évaluer l'influence de cette erreur d'estimation sur les TEBs de la chaîne de tatouage. Nous avons donc procédé à la simulation suivante : les 20 signaux de la base de test ont été tatoués par une séquence de 5000 bits chacun. Le dictionnaire d'émission est constitué de 2 vecteurs antipodaux étalés dans la bande de fréquence $[0; \frac{F_e}{2}]$ et le débit est fixé à 71 bit/s. Le modèle psychoacoustique est celui de Gomes. Quatre modes de réception ont été testés, supposant pour l'occasion que le signal original $x(n)$ est connu du récepteur. Ces modes de réception sont obtenus en faisant varier les combinaisons possibles entre $H(f)$, $\hat{H}(f)$, $G(f)$ et $\hat{G}(f)$ pour le filtrage du signal reçu $\hat{y}(n)$ et le calcul du dictionnaire de réception. Ces cas de figure seront notés HG, $\hat{H}G$, $H\hat{G}$ et $\hat{H}\hat{G}$ au regard des filtres utilisés par le récepteur [2]. Le TEB

2. Rappelons que le cas HG correspond au cas le plus favorable (puisqu'il n'y a pas d'erreur d'estimation sur les filtres) et que le cas $\hat{H}\hat{G}$ est le cas le plus défavorable mais qui sera par la suite mis effectivement en œuvre (puisqu'il permet une détection aveugle).

Mode de réception	$\mathbf{HG}$	$\hat{\mathbf{H}}\mathbf{G}$	$\mathbf{H}\hat{\mathbf{G}}$	$\hat{\mathbf{H}}\hat{\mathbf{G}}$
TEB Sans perturbation	$4,2.10^{-3}$	$3,4.10^{-3}$	$\mathbf{10,9.10^{-3}}$	$8,6.10^{-3}$
TEB Compression MPEG 96 kbit/s	$4,9.10^{-3}$	$9,7.10^{-3}$	$13,5.10^{-3}$	$28,6.10^{-3}$
Rapport Compression/Sans perturbation	$1,16$	$\mathbf{2,8}$	$1,24$	$3,3$

Table 6.2 – Influence de l'erreur d'estimation des filtres de la chaîne de tatouage dans le cas d'un canal sans perturbation et d'une compression MPEG à 96 kbit/s (débit 71 bit/s).

	$\mathbf{DQ(H)}$	$\mathbf{DQ(G)}$
Sans perturbation	-18 dB	-28 dB
Compression MPEG 96 kbit/s	16 dB	-17 dB

Table 6.3 – Estimation de l'erreur d'estimation des filtres de la chaîne de tatouage par la mesure de la déviation quadratique.

moyen dans les quatre cas de figure a été évalué pour deux types de configuration du canal : le cas d'un canal sans perturbation, pour lequel seul l'ajout du signal de tatouage peut introduire une erreur d'estimation des filtres et le cas d'une compression MPEG à 96 kbit/s. La table 6.2 présente les résultats obtenus et indique le rapport entre le TEB avec compression et le TEB sans perturbation. Nous proposons également d'évaluer l'erreur d'estimation par un critère de Déviation Quadratique (DQ) défini directement à partir des coefficients des filtres. Dans le cas du filtre blanchissant $G(f)$, ce critère s'exprime sous la forme :

$$\mathrm{DQ}(G) = \frac{(\mathbf{g} - \hat{\mathbf{g}})^t(\mathbf{g} - \hat{\mathbf{g}})}{\mathbf{g}^t\mathbf{g}}, \tag{6.1}$$

où $\mathbf{g}$ (respectivement $\hat{\mathbf{g}}$) est le vecteur des coefficients du filtre $G(f)$ (respectivement $\hat{G}(f)$). Dans le cas du filtre de mise en forme perceptuelle $H(f)$, il prend l'allure :

$$\mathrm{DQ}(H) = \frac{(\mathbf{a}' - \hat{\mathbf{a}}')^t(\mathbf{a} - \hat{\mathbf{a}}')}{\mathbf{a}'^{\,t}\,\mathbf{a}'}, \text{ avec } \mathbf{a}' = \frac{1}{b_0}\mathbf{a} \text{ et } \hat{\mathbf{a}}' = \frac{1}{\hat{b}_0}\hat{\mathbf{a}}, \tag{6.2}$$

où $\mathbf{a}$, b_0, $\hat{\mathbf{a}}$ et $\hat{b}_0$ désignent successivement le vecteur des coefficients et le gain des filtres $H(f)$ et $\hat{H}(f)$. Les valeurs moyennes des DQ obtenues sur l'ensemble des 20 signaux tests et pour les deux types de canaux sont présentées table 6.3.

Ces expérimentations permettent d'établir les conclusions suivantes. Lorsque le canal ne subit aucune perturbation, l'erreur d'estimation des filtres est relativement faible, dans la mesure où les valeurs de la DQ sont de l'ordre de -18 dB. Néanmoins, la probabilité d'erreur est multipliée par 2 entre le cas idéal HG où le signal audio serait connu du récepteur et le cas $\hat{H}\hat{G}$ d'un détecteur aveugle. Dans le cas d'une compression MPEG, le TEB croît avec les différentes configurations : les TEBs sont multipliés par environ 6 entre le cas idéal HG et le cas pratique $\hat{H}\hat{G}$. La dégradation des TEBs entre le cas d'une compression et le cas sans perturbation révèle l'influence de l'erreur d'estimation du filtre $H(f)$. En effet, les TEBs du cas $\hat{H}G$ sont multipliés par 3 si l'on introduit une compression MPEG. La mesure de la DQ confirme qu'en cas de compression, $H(f)$ et $\hat{H}(f)$ sont très différents. Cette différence entraîne une forte dégradation des TEBs. Il paraît alors nécessaire de choisir un **modèle psychoacoustique robuste**, c'est à dire qui garantisse une faible erreur d'estimation entre $H(f)$ et $\hat{H}(f)$ quelle que soit la perturbation. Cette étude fera l'objet de la section 6.3. Il convient également de s'attarder sur les raisons

de cette erreur d'estimation dans le cas de la compression MPEG. Elles mettent en évidence la nécessité de choisir un signal de tatouage et donc des vecteurs du dictionnaire d'émission à bande limitée $[0; F_c]$. Cette étude fait l'objet de la section suivante.

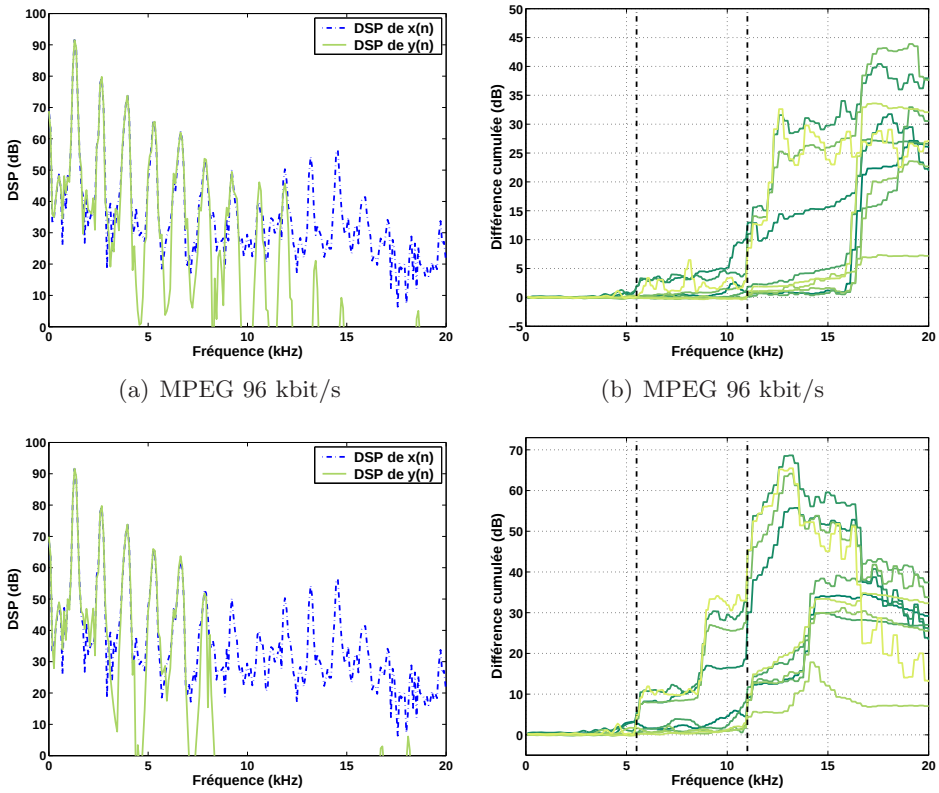
6.2 Réglage de la bande passante : impact de la compression MPEG

Modéliser les effets de la compression MPEG s'avère délicat, puisque le procédé de quantification dépend à la fois du nombre de bits disponibles pour la quantification et du signal audio (via le seuil de masquage qui lui est associé).

L'effet du codeur est couramment assimilé à une opération de filtrage qui supprime les composantes fréquentielles du signal audio jugées non significatives d'un point de vue perceptuel. Cette modélisation est particulièrement adaptée au cas des hautes fréquences. En effet, les hautes fréquences du signal audio sont en grande majorité inférieures au seuil de masquage (qui approche ici le seuil d'audition absolu). Lors du procédé d'allocation binaire, peu de bits sont alloués à la quantification de ces hautes fréquences "presque" inaudibles. L'information qu'elles portent est irrémédiablement perdue. Ce phénomène est mis en évidence figure 6.3. Nous y présentons une comparaison des DSP du signal audio original $x(n)$ et du signal audio comprimé $y(n)$ pour différentes valeurs du débit de compression (96 et 64 kbit/s). Ces DSP sont calculées sur une trame de 512 échantillons d'un signal de violon. Si les basses fréquences du signal sont relativement bien conservées après compression, les hautes fréquences sont en grande partie éliminées. La compression peut donc être associée à un filtre passe-bas de fréquence de coupure F_c. La valeur de F_c ne peut pas être déterminée précisément au regard de ces courbes, d'autant plus qu'elle dépend manifestement du débit de compression : pour 96 kbit/s elle avoisine 12 kHz ; pour 64 kbit/s, elle s'approche de 9 kHz.

Si la compression est modélisable par un filtre passe-bas, le système de tatouage devient un problème de communication à bande limitée $[0; F_c]$. La théorie des communications [Pro01] indique alors que le signal porteur de l'information de tatouage (c'est à dire le signal de tatouage, aussi bien que le signal modulé) doit être choisi dans la bande $[0; F_c]$: en effet pourquoi transmettre une information au delà de la fréquence F_c si l'on sait qu'elle sera perdue après compression ? Cox dans [CKTLS97] aboutit à la même conclusion. Puisque les régions fréquentielles non perceptuelles sont éliminées par la compression MPEG, le tatouage, pour être robuste, doit se situer dans les régions fréquentielles perceptuellement significatives. Ces composantes sont celles qui ne seront pas éliminées par la compression MPEG. Au regard des DSP de la figure 6.3, il s'agit principalement des composantes basses-fréquences. Ainsi les vecteurs du dictionnaire d'émission sont choisis étalés en fréquence, mais limités à la bande de fréquence $[0; F_c]$. Le signal modulé résultant n'est mis en forme que dans la bande de fréquence $[0; F_c]$. Un filtrage Passe-Bas (PB) de fréquence de coupure F_c est donc combiné au calcul du filtre de mise en forme $H(f)$ de sorte que les composantes fréquentielles du signal modulé (introduites notamment par la concaténation des vecteurs du dictionnaire) ne soient pas amplifiées. Ce filtrage passe-bas est également introduit en amont du récepteur : il permet de supprimer du signal reçu les composantes hautes-fréquences (provenant par exemple du signal audio) non significatives pour la réception du tatouage. Le schéma du système de tatouage est donc modifié comme présenté figure 6.4.

Le choix de la fréquence de coupure F_c dépend de la DSP du signal comprimé mais également



(a) MPEG 96 kbit/s (b) MPEG 96 kbit/s

(c) MPEG 64 kbit/s (d) MPEG 64 kbit/s

Figure 6.3 – Impact de la compression MPEG sur la DSP des signaux audio et sur les seuils de masquage pour deux débits de compression (96 et 64 kbit/s) : (a), (c) Comparaison de la DSP d'un signal audio de violon $x(n)$ et celle de sa version comprimée $y(n)$; (b) , (d) Erreur d'estimation moyenne $E_M(f)$ des seuils de masquage (entre le signal audio et sa version comprimée) pour les 10 premiers signaux de la base test (modèle psychoacoustique de Gomes).

du seuil de masquage. En effet, la suppression des hautes fréquences du signal audio tatoué comprimé modifie l'estimation du seuil de masquage et donc du filtre de mise en forme à la réception. L'erreur d'estimation entre ces filtres à l'émission et à la réception entraîne d'après la section précédente une forte augmentation de la probabilité d'erreur. En introduisant un filtre passe-bas dans la chaîne de tatouage, nous limitons la concordance de la réponse du filtre de mise en forme perceptuelle et du seuil de masquage à la bande de fréquence $[0; F_c]$. Pour garantir la robustesse de l'estimation du filtre à l'émission et à la réception, il suffit donc de choisir la fréquence F_c de sorte que l'erreur d'estimation entre les seuils de masquage à l'émission et à la réception dans la bande $[0; F_c]$ soit faible. Plus concrètement, notons $S_M^{(x)}(f, m)$ le seuil de masquage calculé sur la n-ième trame du signal audio original $x(n)$ et $S_M^{(y)}(f, m)$ le seuil de masquage associé à la m-ième trame du signal $y(n)$, résultat de la compression MPEG de $x(n)$. L'erreur d'estimation entre les seuils de masquage sur la m-ième trame peut être évaluée par la différence $S_M^{(x)}(f, m) - S_M^{(y)}(f, m)$. Les seuils de masquage devant être recalculés régulièrement, l'erreur d'estimation moyenne, notée $E_M(f)$, peut donc être évaluée comme la moyenne fréquence

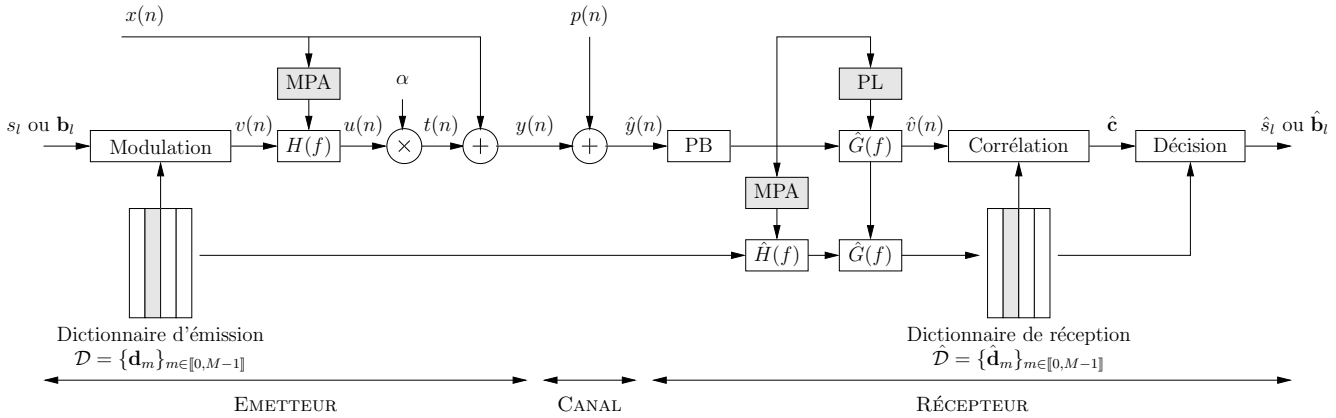


Figure 6.4 – Schéma du système de tatouage avec introduction du filtrage passe-bas.

par fréquence des erreurs d'estimation entre les seuils sur M trames :

$$E_M(f) = \frac{1}{M}\sum_{n=0}^{M-1}\left[S_M^{(x)}(f,m) - S_M^{(y,m)}(f)\right]. \tag{6.3}$$

La figure 6.3 présente les erreurs d'estimation moyenne obtenues après analyse des 10 premiers signaux de la base test sur 5 s de signal et compression pour deux valeurs de débit (96 et 64 kbit/s). Les seuils de masquage ont été calculés avec le modèle psychoacoustique de Gomes, présenté section 5.1.1.2. Ces figures mettent nettement en évidence que l'erreur d'estimation entre les seuils de masquage (bien qu'elle dépende du signal audio) augmentent avec la fréquence et le débit de compression. Des paliers dans l'augmentation de l'erreur apparaissent à $5,5$ kHz et à 11 kHz. Les seuils de masquage diffèrent de moins de 5 dB dans la bande $[0; 5,5]$ kHz si le débit de compression est supérieur à 64 kbit/s. Si la compression MPEG est limitée à 96 kbit/s, cette bande peut être étendue jusqu'à 11 kHz tout en conservant une erreur d'estimation inférieure à 10 dB. Pour prendre en compte tous les débits de compression possibles et limiter l'erreur d'estimation des filtres entre l'émission et la réception, nous choisirons donc la fréquence de coupure $F_c = 5,5$ kHz.

Pour confirmer l'intérêt de limiter la bande de fréquence des vecteurs du dictionnaire du signal de tatouage sur les performances du système, nous présentons table 6.4 les TEBs obtenus par tatouage des 20 signaux de la base test à un débit de 71 bit/s pour deux valeurs de la fréquence de coupure : $F_c = 5,5$ et 11 kHz. Ces TEBs sont mesurés dans le cas d'un canal sans perturbation et dans le cas d'une compression MPEG à différents débits (96 et 64 kbit/s). Nous présentons également table 6.5 les mesures de déviation quadratique moyenne entre les filtres de mise en forme perceptuelle $H(f)$ calculé à l'émetteur et $\hat{H}(f)$ calculé au récepteur pour les mêmes configurations de bande passante et de canal que celles décrites précédemment. Les résultats obtenus pour $F_c = \dfrac{F_e}{2}$ dans le cadre de l'évaluation de l'influence d'erreur d'estimation des filtres (table 6.2 et 6.5) y ont été reportés. Les résultats confirment que diminuer la largeur de bande F_c conduit à réduire l'erreur d'estimation du filtre $H(f)$: cette erreur est inférieure à -20 dB lorsque $F_c = 5,5$ kHz même en présence de compression (dont les débits sont supérieurs à 64 kbit/s). Les TEBs de la chaîne sont directement liés à l'erreur d'estimation : plus cette erreur devient petite et meilleur est le TEB. En choisissant $F_c = 5,5$ kHz, le système de tatouage

Canal	Sans perturbation	MPEG 96 kbit/s	MPEG 64 kbit/s
$F_c = F_e/2$	$8,6.10^{-3}$	$28,6.10^{-3}$	
$F_c = 11$ kHz	3.10^{-3}	$3,6.10^{-3}$	$5,7.10^{-3}$
$F_c = 5,5$ kHz	$1,6.10^{-3}$	2.10^{-3}	$2,6.10^{-3}$

Table 6.4 – Influence de la largeur de bande F_c sur les TEBs de la chaîne de tatouage dans le cas d'un canal sans perturbation et avec compression MPEG à 96 et 64 kbit/s (débit 71 bit/s).

Canal	Sans perturbation	MPEG 96 kbit/s	MPEG 64 kbit/s
$F_c = F_e/2$	-18 dB	**16 dB**	
$F_c = 11$ kHz	-22 dB	-20 dB	-11 dB
$F_c = 5,5$ kHz	-24 dB	-24 dB	-23 dB

Table 6.5 – Mesure de l'erreur d'estimation DQ(H) du filtre de mise en forme en fonction de la largeur de bande F_c dans le cas d'un canal sans perturbation et avec compression MPEG à 96, 64 et 32 kbit/s (débit 71 bit/s).

devient robuste à la compression MPEG. En effet, bien que l'erreur d'estimation du filtre soit importante (la déviation quadratique étant de 34 dB), le TEB obtenu à ce débit est faible, d'un même ordre de grandeur que celui du cas sans perturbation. Par la suite nous travaillerons donc avec une bande limitée à $[0; 5,5]$ kHz.

6.3 Seuil de masquage

6.3.1 Critères de choix

Le seuil de masquage joue un rôle majeur dans la chaîne de tatouage : il garantit l'inaudibilité du signal de tatouage en présence du signal audio tout en permettant l'augmentation de la puissance du signal de tatouage par rapport à celle du signal audio, ce qui facilite sa détection. Le choix du seuil de masquage relève donc d'un compromis entre l'inaudibilité et le rapport de puissance entre tatouage et audio. A transparence du signal de tatouage équivalente, on aura donc tendance à choisir un seuil de masquage qui permet un rapport de puissance maximal. Le signal de tatouage étant construit pour que sa DSP coïncide avec le seuil de masquage à un facteur multiplicatif α près, ce rapport peut être évalué par le RMS, défini par :

$$\text{RMS} = \frac{\int_0^{F_c} S_M(f)df}{\int_0^{F_c} S_X(f)df}, \tag{6.4}$$

où $S_M(f)$ désigne le seuil de masquage et $S_X(f)$ désigne la DSP du signal audio $x(n)$. Ce rapport masque à signal est également un critère fréquemment utilisé pour caractériser la transparence du tatouage : plus le RMS est faible et plus le tatouage est inaudible.

Le seuil de masquage doit également satisfaire à une contrainte de robustesse liée au filtre de mise en forme perceptuelle. La section 6.1.2 montre que les TEBs sont directement liés à la différence d'estimation entre le filtre de mise en forme calculé à l'émetteur et sa version réestimée à la réception : plus ces deux filtres sont semblables et meilleur est le taux d'erreur binaire. Cette conclusion se répercute directement sur les seuils de masquage, puisque les filtres sont construits

de sorte à ce que leur réponse fréquentielle coïncide avec le seuil. Plus les seuils de masquage obtenus à partir du signal audio original et du signal audio tatoué sont similaires, meilleure en est la réception. Il convient donc de choisir un modèle psychoacoustique robuste, c'est à dire un modèle pour lequel le seuil de masquage $\hat{S}_M(f)$ calculé à partir du signal audio tatoué détérioré $\hat{y}(n)$ soit le plus similaire au seuil de masquage $S_M(f)$ du signal audio original :

$$\forall f, \hat{S}_M(f) \approx S_M(f). \tag{6.5}$$

Quatre critères interviennent finalement dans le choix d'un modèle psychoacoustique adapté : la transparence du tatouage, le RMS, la probabilité d'erreur et la robustesse. Nous pouvons également y ajouter un critère de coût en temps de calcul, qui peut faire pencher la balance pour des applications temps-réel.

6.3.2 Performances des seuils de masquage

Dans cette section, nous nous proposons d'évaluer les performances des trois modèles psychoacoustiques évoqués section 5.1.1.2 : le MPA n°1 de MPEG, le modèle proposé par Gomes [Gom02] et celui de Garcia [Gar99]. Cette comparaison ne peut être effectuée qu'à la condition que ces modèles engendrent des conditions de transmission identiques : on souhaite notamment que les signaux de tatouage résultant de la mise en forme perceptuelle par ces modèles soient "identiquement inaudibles". Pour juger de cette inaudibilité, nous avons choisi d'utiliser le critère RMS (bien que ses performances soient contestables). En imposant une valeur de RMS, nous garantissons également que les signaux de tatouage obtenus par les différents modèles sont "identiquement puissants" et offrent des conditions de détection en matière de rapport signal à bruit relativement identiques. Chaque modèle est donc paramétré de sorte que le RMS soit constant. Pour ce faire, nous pouvons agir sur le facteur d'amplitude α qui pondère le signal de tatouage.

Cette précaution étant prise, nous pouvons comparer les performances des modèles. L'inaudibilité étant supposée identique pour les trois modèles en imposant le RMS à -25 dB, trois critères ont été mesurés [Bat05] : le TEB, la robustesse du seuil et le temps de calcul moyen. Le TEB est obtenu en tatouant les 20 signaux de la base de test avec 5000 bits chacun, pour un dictionnaire de 2 vecteurs antipodaux étalés dans la bande $[0; 5,5]$ kHz, et un débit de 71 bit/s. Ce TEB est mesuré dans le cas d'un canal sans perturbation et dans le cas d'une compression MPEG à 96 kbit/s. La robustesse des seuils est mesurée via le critère de déviation quadratique entre le filtre de mise en forme $H(f)$ calculé à l'émission et à la réception. Le temps de calcul est le temps moyen de calcul du seuil de masquage pour un PC dont les caractéristiques ont été précisées précédemment, section 2.4. Les résultats obtenus sont présentés table 6.6. En ce qui concerne les TEBs (quelle que soit la perturbation), le modèle MPEG fournit les meilleures performances, suivi de peu par le modèle de Gomes [3]. Par contre, le modèle de Garcia est peu efficace (puisque les valeurs obtenues sont 10 fois supérieures aux TEBs du modèle MPEG). Cette tendance s'inverse lorsqu'il est question de la robustesse du seuil : l'erreur d'estimation sur les filtres de mise en forme perceptuelle est très faible dans le cas du modèle de Garcia et augmente avec le modèle MPEG. Le modèle de Garcia est donc le plus robuste des 3 modèles psychoacoustiques considérés [4]. Enfin, le temps de calcul montre que le modèle de Gomes et le

3. L'ordre de grandeur des TEBs obtenus est, ici, plus élevé que dans les simulations précédentes. Ceci est du au choix du RMS à -25 dB, qui est nettement moins favorable à la transmission du tatouage que dans les simulations précédentes

4. En effet, le système de Garcia exploite le seuil de masquage pour déterminer les régions fréquentielles non

MPA		MPEG	Gomes	Garcia
TEB	Sans perturbation	$\mathbf{1,7.10^{-2}}$	3.10^{-2}	$14,5.10^{-2}$
	MPEG 96 kbit/s	$\mathbf{1,9.10^{-2}}$	$3,2.10^{-2}$	$15,3.10^{-2}$
DQ(H)	Sans perturbation	-20 dB	-24 dB	**-30 dB**
Temps de calcul		10 ms	**0,4 ms**	0,9 ms

Table 6.6 – Performances de trois modèles psychoacoustiques en terme de TEB (sans perturbation et avec compression MPEG à 96 kbit/s), de robustesse du seuil de masquage (par la mesure de la déviation quadratique du filtre de mise en forme $H(f)$) et de temps de calcul (RMS = -25 dB).

modèle de Garcia sont peu coûteux. Par contre, le modèle MPEG pâtit d'un temps de calcul très important avec 10 ms pour traiter chaque seuil.

Chaque modèle bénéficie donc d'avantages et d'inconvénients différents : le modèle MPEG est performant pour le TEB mais très coûteux, tandis que le modèle de Garcia est robuste mais induit des TEBs élevés. Le modèle de Gomes paraît offrir le meilleur compromis entre les 3 critères mesurés : des TEBs acceptables (quelle que soit la perturbation), une robustesse correcte et un temps de calcul très faible. Par la suite, nous ne conserverons que le MPA de Gomes.

6.4 Contrôle de l'inaudibilité par un facteur d'amplitude α adapté au signal audio

6.4.1 Nécessité d'un facteur d'amplitude adaptatif

Le seuil de masquage ayant été choisi, nous nous intéressons maintenant au choix du facteur d'amplitude α. Ce paramètre permet de régler le niveau de transparence du tatouage. Dans la section précédente, nous avions choisi de contrôler l'inaudibilité via le RMS. Ce rapport ne fournit malheureusement qu'une mesure globale de la distorsion introduite par le tatouage et ne permet pas de prendre en compte d'éventuelles distorsions locales introduites à des fréquences spécifiques. Il convient alors de faire appel à l'algorithme PEAQ décrit section 2.1.3. Cet algorithme fournit une note, l'ODG, qui caractérise la qualité du signal audio tatoué $y(n) = x(n) + \alpha u(n)$ par rapport au signal audio original $x(n)$: cette note varie entre 0 (lorsque le tatouage est transparent) à -4 (dans le cas où la dégradation est très importante). Le niveau de transparence du tatouage est lié au facteur d'amplitude α qui paramétrise le mélange entre le signal audio $x(n)$ et le signal modulé après mise en forme spectrale $u(n)$.

ODG en fonction de α

La figure 6.5 présente les valeurs moyennes de l'ODG et les RMS correspondants en fonction du facteur d'amplitude α. Ces valeurs sont obtenues par tatouage des 20 signaux de la base de test, à un débit de 71 bit/s sur 15 s de signal chacun. Ces courbes confirment que la transparence

significatives du signal audio (celles pour lesquelles le seuil de masquage est supérieur à la DSP du signal audio). Ces régions fréquentielles seront remplacées par l'information de tatouage. Pour permettre une détection aveugle du tatouage, elles doivent donc être réestimées à partir du signal audio tatoué de manière suffisamment précise pour chercher le tatouage au bon endroit. L'efficacité du système de Garcia est donc principalement basée sur la robustesse de son seuil.

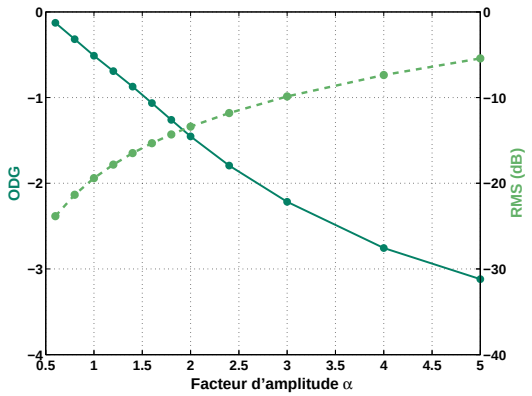


Figure 6.5 – Variation de l'ODG moyenne (courbe en trait plein) et du rapport masque à signal RMS (courbe en pointillé) en fonction du facteur d'amplitude α sur le corpus de 20 signaux.

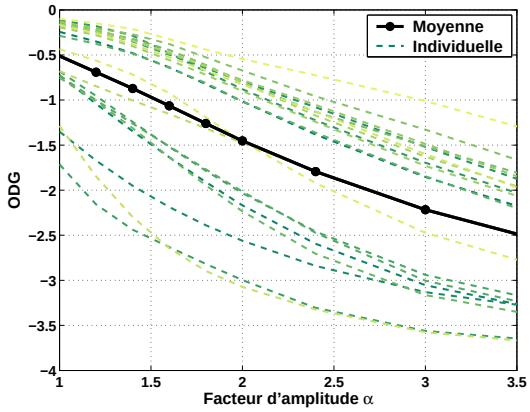


Figure 6.6 – ODG en fonction de α : la courbe en trait plein rappelle l'ODG moyenne obtenue sur le corpus et les 20 courbes en pointillé l'ODG individuelle obtenue par chaque signal.

du tatouage est proportionnelle au facteur d'amplitude α et est inversement proportionnelle au RMS. Elles permettent surtout de choisir pour une qualité d'inaudibilité donnée une valeur de α adaptée : notamment, si l'on souhaite obtenir un tatouage perceptible mais non gênant (pour lequel l'ODG est de -1), nous serions enclins à choisir la valeur $\alpha = 1,5$ permettant d'atteindre en moyenne cette ODG.

Limites du modèle psychoacoustique

Nous nous intéressons maintenant non plus à la valeur moyenne de l'ODG sur les 20 signaux du corpus mais à l'ODG individuelle obtenu par chaque signal en fonction de α. Les courbes présentées figure 6.6 montrent l'évolution de l'ODG en fonction de α pour chacun des 20 signaux. Elles mettent en évidence que l'inaudibilité obtenue (i.e. l'ODG) est fortement dépendante du signal audio. Pour une valeur de α fixée, par exemple à $1,5$ si l'on souhaite obtenir une ODG de -1, une forte variabilité de l'ODG individuelle est observée : pour certains signaux, l'ODG est supérieure à la valeur souhaitée, induisant un tatouage "plus" transparent que ce que nous souhaitions obtenir. Inversement pour d'autres signaux, l'ODG est nettement plus faible que la valeur attendue -1. Dans ces derniers cas, le tatouage est nettement plus audible qu'escompté : deux cas de figure[5] atteignent notamment une ODG de $-2,5$, traduisant un tatouage audible et gênant.

Cette étude met en évidence la nécessité d'adapter α non pas globalement (c'est à dire à partir d'une valeur moyenne de l'ODG) mais directement en fonction de chaque signal audio, voire même sur chaque trame de signal audio. Tel est l'objectif de l'algorithme présenté par la suite.

5. Il s'agit des signaux *"vivaldi"* et *"cranberries"* : l'un est un extrait d'une symphonie par un orchestre, l'autre un extrait d'une chanson rock. Cette diversité ne permet donc pas d'établir un lien entre la nature du signal et l'ODG obtenue.

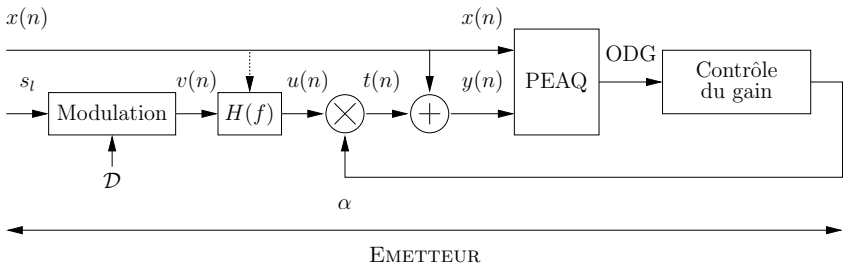


Figure 6.7 – Système de tatouage audio avec module du contrôle de l'inaudibilité permettant le choix du gain α adapté.

6.4.2 Algorithme de contrôle adaptatif de l'inaudibilité

Pour adapter localement les paramètres d'inaudibilité en fonction du signal audio, et donc choisir un facteur de gain α adapté, l'algorithme PEAQ est introduit à l'émetteur. Cet algorithme permet de mesurer tout au long de la construction du signal audio tatoué $y(n) = x(n) + t(n)$ avec $t(n) = \alpha u(n)$ l'ODG atteinte puis de rétroagir sur le choix du gain α de sorte à garantir une valeur de l'ODG donnée. Cette valeur d'inaudibilité à atteindre, paramétrable, sera notée ODG_a. Le schéma de l'émetteur ainsi obtenu est présenté figure 6.7.

Supposons pour l'instant que la construction du signal modulé $u(n)$ mis en forme perceptuellement est indépendante du facteur d'amplitude α. Cette hypothèse est valable pour le système de tatouage construit préalablement et le reste tant que la modulation n'est pas dépendante du signal audio.

Notons $\mathbf{u}$ le signal modulé mis en forme perceptuellement, construit pour une séquence binaire donnée et observé sur une trame de durée N_m ; de façon équivalente $\mathbf{x}$ est le signal audio, dans lequel $\mathbf{u}$ doit être inséré, observé sur la même trame de durée N_m. L'objectif de la stratégie d'insertion est alors de déterminer la valeur adaptée du gain α qui garantit que la valeur de l'ODG obtenue en comparant le signal audio $\mathbf{x}$ et le signal audio tatoué $\mathbf{y} = \mathbf{x} + \alpha\mathbf{u}$ concorde avec la valeur à atteindre ODG_a à une erreur ΔODG près.

L'analyse de l'évolution de l'ODG en fonction du gain α (donnée par les courbes 6.6) montre que l'ODG est une fonction décroissante de α. Ainsi un algorithme itératif très simple peut être mis en œuvre pour faire converger l'ODG vers ODG_a en jouant sur le paramètre α. Pour faciliter la convergence de l'algorithme, la progression de α est dichotomonique [6]. L'algorithme peut donc être formalisé par les étapes décrites dans l'algorithme 6.1. En pratique, cet algorithme converge en quelques itérations (moins d'une dizaine).

Deux paramètres sont à fixer : la taille des fenêtres d'analyse N_m et la précision de convergence de l'algorithme ΔODG. Le choix de ces paramètres est lié à la conception même de l'algorithme PEAQ. En effet, cette méthode fournit une mesure de la qualité sonore en moyennant un ensemble de caractéristiques représentant les différences de perception entre les deux signaux. La norme, qui décrit la méthode, établit que cette mesure est significative si les signaux ont une durée d'environ 10 s et que sa précision est de $0,1$. La convergence de l'algorithme est donc naturellement choisie égale à $\Delta\text{ODG} = 0,1$. La taille des fenêtres ne doit pas être trop faible (elle ne peut notamment pas correspondre à la longueur $N_H = 512$ échantillons d'actualisation

6. c'est à dire que sa recherche est réalisée dans un ensemble ordonné, partagé en deux à chaque itération, pour augmenter la rapidité de convergence de l'algorithme.

Étape 0 : **Initialisations**

- α est initialisé par une valeur α_0, choisie en fonction de la valeur de l'ODG à atteindre ODG_a. Cette valeur est obtenue en utilisant la courbe 6.6 donnant l'ODG moyenne des 20 signaux du corpus en fonction de α. Par exemple, si $\text{ODG}_a = -1$, on choisit naturellement $\alpha^{(0)} = 1,5$.
- l'intervalle de recherche $\mathbb{A} = [\alpha^{(inf)}, \alpha^{(sup)}]$ de α est initialisé à $[0, \alpha^{(0)}]$.

Étape 1 : **Calcul de l'ODG pour la valeur de $\alpha^{(n)}$ en cours**

L'algorithme PEAQ est mis en œuvre pour évaluer la valeur de l'ODG, noté ODG, caractérisant la distorsion perceptuelle entre le signal audio $\mathbf{x}$ et sa version tatouée $\mathbf{y} = \mathbf{x} + \alpha\mathbf{u}$.

Étape 2 : **Actualisation dichotomique de $\alpha^{(n)}$**

L'ODG obtenue est comparée à la valeur à atteindre :

- si $|\text{ODG} - \text{ODG}_a| < \Delta\text{ODG}$, l'algorithme s'arrête.
- sinon si $\text{ODG} > \text{ODG}_a$ (resp. $\text{ODG} < \text{ODG}_a$), α est décrémenté (resp. incrémenté) de manière dichotomique et l'étape 1 est alors réitérée. Cette actualisation dichotomique s'effectue de la façon suivante : étant donné l'intervalle de recherche $\mathbb{A} = [\alpha^{(inf)}, \alpha^{(sup)}]$,
  - si $\text{ODG} > \text{ODG}_a$, $\alpha^{(n+1)} = \alpha^{(inf)} + \dfrac{\alpha^{(n)} - \alpha^{(inf)}}{2}$ et l'intervalle de recherche devient $\mathbb{A} = [\alpha^{(inf)}, \alpha^{(n)}]$.
  - si $\text{ODG} < \text{ODG}_a$, alors,
    - si $\alpha^{(n)} = \alpha^{(sup)}$, alors de fait, $\alpha^{(n+1)} = 2\alpha^{(n)}$ et $\mathbb{A} = [\alpha^{(n)}, \alpha^{(n+1)}]$,
    - sinon, $\alpha^{(n+1)} = \alpha^{(n)} + \dfrac{\alpha^{(sup)} - \alpha^{(n)}}{2}$ et $\mathbb{A} = [\alpha^{(n)}, \alpha^{(sup)}]$.

Algorithme 6.1: Contrôle adaptatif de l'inaudibilité

du filtre de mise en forme) mais ne doit pas être trop élevée non plus pour permettre une actualisation locale du gain α. Nous avons choisi des fenêtres d'une seconde ; ce choix correspond à la structure du signal établie pour la synchronisation (qui sera présentée section 11), ce qui permet une implémentation aisée du système de tatouage. Pour chaque fenêtre d'analyse, l'algorithme conserve les caractéristiques de perception des 9 secondes de signal précédant la fenêtre d'analyse et les utilisent pour la mesure de l'ODG de la fenêtre en cours. Après calcul du gain α adapté à la fenêtre en cours, les 9 secondes de signal les plus récentes sont conservées pour l'analyse de la fenêtre suivante.

Le choix de ce fenêtrage relativement court présente un inconvénient : dans certains configurations de signal, une augmentation brutale du gain α peut se produire induisant l'introduction d'une dégradation auditive locale. Pour éviter ce phénomène une fenêtre d'oubli exponentielle est introduite pour ralentir les variations de α.

6.4.3 Performances de l'algorithme

Pour évaluer l'efficacité de cette stratégie de contrôle d'inaudibilité, nous étudions figure 6.8 les variations du gain α et de l'ODG en fonction du temps pour deux signaux du corpus :

"Partita_no3" et *"Gainsbourg"* : les figures (a) et (b) présentent l'évolution temporelle des signaux audio originaux $x(n)$ et du tatouage résultant $\alpha u(n)$; les figures (c) et (d) exposent l'évolution temporelle du gain α, tandis que les figures (e) et (f) s'attachent à la valeur de l'ODG résultante en fonction du temps.

Les deux signaux étudiés ont été spécifiquement choisis dans le corpus parce qu'ils présentent deux valeurs extrêmes de l'ODG lorsque $\alpha = 1,5$. La paramétrisation du système avait été effectuée de sorte que l'ODG moyenne obtenue pour les 20 signaux du corpus soit -1 (traduisant un tatouage audible mais non gênant). L'ODG obtenue par le signal *"Partita_no3"*, $-2,6$, est nettement inférieure, ce qui signifie que l'insertion du tatouage introduit donc une dégradation audible et gênante, tandis que celle du signal *"Gainsbourg"*, $-0,3$, est supérieure à -1 (le tatouage est ici inaudible).

L'ODG à atteindre ODG_a par l'algorithme de contrôle d'inaudibilité est là encore fixée à -1 (pour une précision de $\Delta\text{ODG} = 0,1$). Les figures mettent en évidence l'efficacité de l'algorithme puisque l'ODG des deux signaux varie à proximité de -1. Dans le cas du signal *"Partita_no3"* très sensible à l'introduction du tatouage, le gain α adapté varie autour de 1 (et est donc effectivement inférieur à la valeur $1,5$ qui produit un tatouage audible). A l'inverse dans le cas du signal *"Gainsbourg"*, peu sensible à l'ajout du tatouage, le gain α adapté est nettement supérieur à $1,5$.

Ces courbes mettent également en évidence l'efficacité de la fenêtre d'oubli pour atténuer les augmentations brutales du gain en fonction du signal audio traité (nettement visibles dans le cas du signal *"Gainsbourg"*) sans pour autant nuire à la convergence de l'ODG vers sa valeur attendue $\text{ODG}_a = -1$. Par contre, elle atténue voire retarde la diminution du gain sur des fenêtres d'analyse où le signal audio est très sensible à l'insertion du tatouage. En empêchant cette diminution du gain, nous risquons d'introduire des dégradations auditives sur certaines fenêtres d'analyse. Il faudrait donc modifier cette fenêtre d'oubli pour qu'elle puisse atténuer les augmentations du gain sans modifier ses diminutions. Nous n'avons malheureusement pas eu le temps de réaliser cette étude ; néanmoins les résultats obtenus montrent la faisabilité de ce module de contrôle de l'inaudibilité. Pour l'instant, il faudra être vigilant sur les résultats en matière d'inaudibilité en combinant la mesure de l'ODG à des tests d'écoute.

Plusieurs simulations pour un même signal audio ont été effectuées en tatouant différentes séquences binaires (à l'aide de différents dictionnaires d'émission). Ces simulations montrent que le gain adapté α converge toujours vers une même valeur. Nous pourrons donc considérer par la suite que le choix de α est indépendant du signal modulé $\mathbf{v}$ et de sa mise en forme perceptuelle $u(n)$.

6.4.4 Conséquences sur le récepteur

L'introduction de ce module de contrôle d'inaudibilité dans le système de tatouage est immédiate et ne modifie pas le schéma de réception qui est défini indépendamment du facteur d'amplitude α (sous l'hypothèse que les vecteurs du dictionnaire de réception sont de même norme). Par contre, les variations de α tout au long du signal audio tatoué entraînent la modification du rapport de puissance entre le tatouage $\mathbf{t} = \alpha\mathbf{u}$ et le signal audio. Ces variations du RSB modifient le TEB. L'évaluation des performances du système de tatouage avec contrôle d'inaudibilité s'attache donc aussi bien à la valeur de l'ODG obtenue après tatouage qu'à la fiabilité de transmission. Le protocole expérimental utilisé est identique à celui décrit section 2.4, auquel s'ajoute la valeur $\text{ODG}_a = -1$ de l'ODG attendue à une précision $\Delta\text{ODG} = 0,1$ près.

Paramètres	Descriptions	Valeurs choisies
$\mathbf{N_H}$	Durée des fenêtres d'analyse pour le calcul du modèle psychoacoustique	512 échantillons
$\mathbf{N_G}$	Durée des fenêtres d'analyse pour le calcul des filtres blanchissants	512 échantillons
$\mathbf{P_H}$	Ordre du filtre $H(f)$	50
$\mathbf{P_G}$	Ordre du filtre $G(f)$	200
$\mathbf{F_c}$	Limite d'étalement pour les vecteurs du dictionnaire	$5,5$ kHz
α	Facteur d'amplitude	adaptatif de sorte à atteindre une ODG de -1 avec une précision de $0,1$
$\mathbf{S_M(f)}$	Seuil de masquage	issu du MPA de Gomes

Table 6.7 – Paramétrisation du système de tatouage.

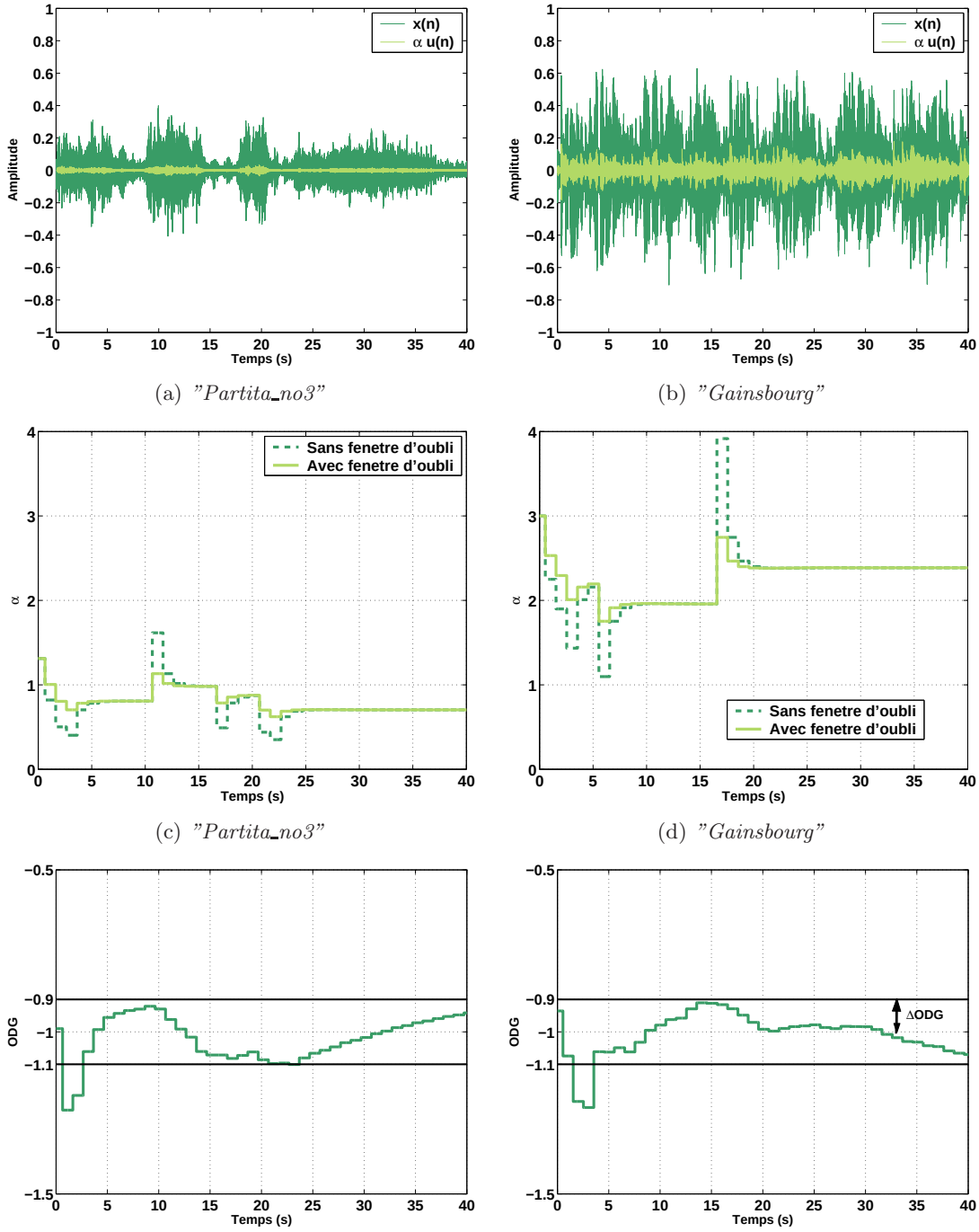
Par la suite, le système sera référé en tant que **système de référence** lorsque le module de contrôle adaptatif du gain α est désactivé et sera dénommé **système de tatouage avec contrôle de l'inaudibilité**, lorsque le module de contrôle est actif.

Conclusion

Ce chapitre fait état d'un effort important de paramétrisation du système de tatouage. Cette paramétrisation influe sur les performances de la chaîne tant en matière de taux d'erreur binaire que de transparence du tatouage. Les différents paramètres considérés et les valeurs choisies sont rappelés table 6.7.

Le choix du modèle psychoacoustique, la mise en œuvre des filtres de la chaîne $H(f)$ et $G(f)$ (établie par l'ordre des filtres P_H et P_G et la taille des fenêtres d'analyse N_H et N_G) et le choix de la bande passante F_c ont notamment été déterminés pour satisfaire aux objectifs de fiabilité de transmission et de robustesse aux perturbations (notamment la compression MPEG).

L'étude menée sur le facteur de gain α, dont dépend la transparence du tatouage, a mis en évidence la forte dépendance des performances de la chaîne au signal audio traité. Un module de contrôle d'inaudibilité, au principe novateur, a été proposé pour choisir spécifiquement le gain α qui garantit un niveau de transparence donnée ODG_e en fonction du signal audio traité. Ce module exploite un algorithme de mesure de la différence perceptuelle entre deux signaux, l'algorithme PEAQ, qui fournit une note, l'ODG, sur une échelle de 0 à -4 (0 correspondant à un tatouage inaudible et -4 à un tatouage très audible). Étant donnée la transparence attendue par l'utilisateur, le module de contrôle rétroagit sur le choix de α pour garantir une transparence proche de celle attendue. Le choix de α devient donc adaptatif.



(a) *"Partita_no3"*

(b) *"Gainsbourg"*

(c) *"Partita_no3"*

(d) *"Gainsbourg"*

(e) *"Partita_no3"*

(f) *"Gainsbourg"*

Figure 6.8 – Évolution du gain α et de l'ODG entre le signal audio $x(n)$ et le signal audio tatoué $y(n) = x(n) + \alpha u(n)$ en fonction du temps pour deux signaux *"Partita_no3"* et *"Gainsbourg"* : (a), (b) variations de $x(n)$ et de $\alpha u(n)$; (c), (d) variations de α et (e), (f) variations de l'ODG.

Chapitre 7

Performances du système de tatouage

La conception du système du tatouage ayant été établie et sa paramétrisation justifiée, ce dernier chapitre est dédié à l'évaluation des performances. Les quatre critères de performances établis section 2.4 sont naturellement considérés : la mesure de la transparence du tatouage, la probabilité d'erreur de transmission en fonction du débit, la robustesse aux perturbations et le coût en temps de calcul de la chaîne de tatouage. La mesure de ces quatre critères sur le corpus de signaux audio établira une référence, servant à évaluer l'apport des stratégies d'insertion informées développées dans la troisième partie.

La mise en œuvre du système sur des signaux réels est également l'occasion d'établir les conditions de transmission effectives de l'information à tatouer et de valider les hypothèses proposées pour la conception du système de tatouage. La première section de ce chapitre s'intéresse donc à vérifier l'optimalité du récepteur : peut-on effectivement justifier de l'orthonormalité du dictionnaire ou d'une IES négligeable ? Elle permet également d'établir les conditions de transmission en terme de rapport signal à bruit qui influe sur la fiabilité de transmission. Ces études permettent de proposer un modèle du système de tatouage sous la forme d'une chaîne de communication BBAG. Les performances théoriques de ce modèle établies section 4.4 pourront finalement être comparées aux performances expérimentales, données dans la seconde section.

7.1 Conditions de transmission

Cette section se propose de valider les hypothèses posées pour la construction du système de tatouage pour justifier de l'optimalité du récepteur. Ces hypothèses sont de deux natures : la première tient à l'orthogonalité et à la puissance des vecteurs du dictionnaire de réception ; la seconde s'attache à l'IES supposée négligeable pour permettre une détection temps symbole par temps symbole.

7.1.1 Etude du dictionnaire de réception

Dans cette étude, nous nous souhaitons évaluer l'orthogonalité et la puissance des vecteurs du dictionnaire de réception. L'orthogonalité du dictionnaire validerait l'optimalité du récepteur par filtre blanchissant et démodulateur par corrélation du système de tatouage tandis que la

puissance de ces vecteurs, si elle est égale, permettrait d'utiliser la corrélation linéaire plutôt que la corrélation normalisée (définies section 3.1.1.4) pour le démodulateur. Ce dictionnaire $\hat{\mathcal{D}} = \left\{\hat{\mathbf{d}}_m\right\}_{m\in[\![0,M-1]\!]}$ résulte du filtrage des vecteurs orthogonaux (ou biorthogonaux), de même puissance unité, du dictionnaire d'émission $\mathcal{D}$. Le filtrage ne conservant pas les propriétés d'orthogonalité, la conservation de l'orthogonalité du dictionnaire à la réception ne peut être garantie. Deux critères d'évaluation sont donc envisagés :

- un ***critère de Conservation de Puissance*** *(CP)*, noté CP. Pour évaluer les différences de puissance entre les vecteurs du dictionnaire de réception, ce critère calcule l'écart relatif moyen entre la puissance des vecteurs et la plus grande valeur de la puissance :

$$\mathrm{CP} = \frac{1}{M-1}\sum_{m\neq g}\frac{\left|\hat{\mathbf{d}}_g^t\hat{\mathbf{d}}_g - \hat{\mathbf{d}}_m^t\hat{\mathbf{d}}_m\right|}{\hat{\mathbf{d}}_g^t\hat{\mathbf{d}}_g} \text{ avec } g = \arg\min_{m\in[\![0,M-1]\!]}\frac{\hat{\mathbf{d}}_m^t\hat{\mathbf{d}}_m}{N_s}, \tag{7.1}$$

  où $|x|$ désigne la valeur absolue de x. Si ce rapport est proche de 0, nous pourrons supposer que les vecteurs du dictionnaire de réception sont tous en moyenne de même puissance.

- un ***critère de Conservation d'Orthogonalité*** *(CO)*, noté CO. L'orthogonalité du dictionnaire de réception n'est plus satisfaite à partir du moment où deux vecteurs du dictionnaire $\hat{\mathbf{d}}_m$ et $\hat{\mathbf{d}}_p$ ont une corrélation non nulle qu'elle soit positive ou négative. L'angle entre ces deux vecteurs, établi par :

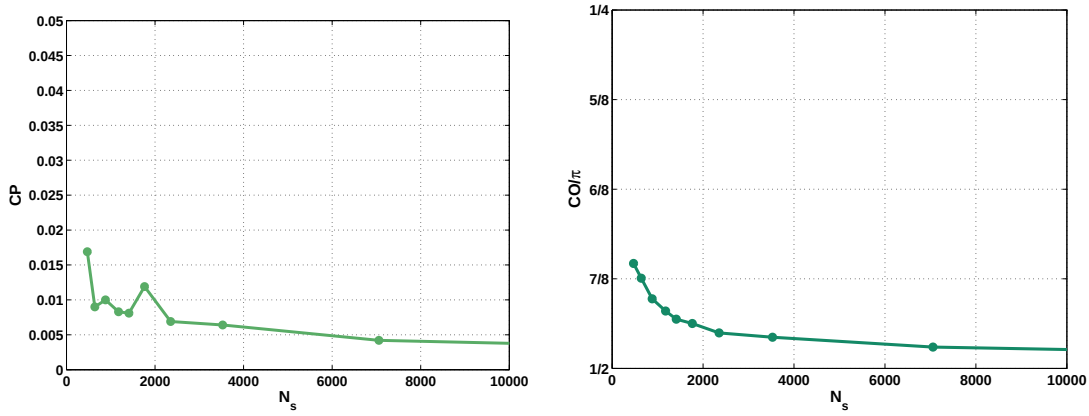
$$\theta = \arccos\left(\frac{\hat{\mathbf{d}}_m^t\hat{\mathbf{d}}_p}{||\hat{\mathbf{d}}_m||\,||\hat{\mathbf{d}}_p||}\right), \tag{7.2}$$

  diffère alors de $\frac{\pi}{2}$. La conservation de l'orthogonalité peut donc être évaluée en mesurant l'angle moyen (sans information de signe) entre chacun des vecteurs du dictionnaire :

$$\mathrm{CO} = \frac{1}{M(M-1)}\sum_{m}\sum_{p\neq m}\arccos\left(\left|\frac{\hat{\mathbf{d}}_m^t}{||\hat{\mathbf{d}}_m||}\frac{\hat{\mathbf{d}}_p}{||\hat{\mathbf{d}}_p||}\right|\right). \tag{7.3}$$

Le dictionnaire de réception est réévalué sur chaque temps symbole pendant lequel s'effectue la détection. Ces caractéristiques varient donc d'un temps symbole à un autre ; elles dépendent à la fois de la durée des vecteurs du dictionnaire d'émission N_s, mais également des filtres de la chaîne, qui dépendent eux-même du signal audio et de la fenêtre sur laquelle il est analysé. On présente donc figure 7.1 les valeurs moyennes des critères d'évaluation CO et CP en fonction de la durée N_s des vecteurs. Ces valeurs sont obtenues après tatouage des 20 signaux du corpus par 5000 bits chacun (soit 5000 calculs de dictionnaire de réception) et un dictionnaire de $M = 8$ vecteurs.

Ces courbes mettent en évidence la perte des propriétés d'orthogonalité et d'égalité en puissance des vecteurs du dictionnaire à la réception. Néanmoins, quel que soit le temps symbole, l'écart relatif moyen des puissances des vecteurs du dictionnaire reste inférieur à 2% : nous pourrons supposer que l'égalité des puissances à la réception est à peu près conservée. Pour des temps symboles relativement élevés (c'est à dire des débits de transmission faibles), l'angle moyen formé par les vecteurs du dictionnaire est proche de $\frac{\pi}{2}$, donc le dictionnaire est presque orthogonal. Cet angle se dégrade lorsque le temps symbole diminue (i.e. le débit augmente), mais reste inférieur à $\frac{7\pi}{8}$: la corrélation entre les vecteurs n'excède donc pas $0,1$. Ainsi, les faibles



(a) (b)

Figure 7.1 – Critère de puissance (a) et critère d'orthogonalité (b) du dictionnaire de réception en fonction du temps symbole N_s obtenus pour un dictionnaire de $M = 8$ vecteurs.

valeurs obtenues par les critères de puissance et d'orthogonalité justifient que nous modélisions (en première approximation) les vecteurs dictionnaire de réception comme orthogonaux et de même puissance.

7.1.2 Interférence entre symboles (IES)

L'IES naît de l'utilisation de filtres à l'émission et à la réception [Bro97]. Lorsque la durée de la réponse impulsionnelle de ces filtres est supérieure au temps symbole N_s, les symboles débordent les uns sur les autres. La détection du symbole sur le l-ième temps symbole est donc perturbée par les symboles qui le précèdent (ou le suivent) ; le récepteur n'est plus optimal.

L'IES peut être vue comme un bruit, qui s'ajoute au bruit de la chaîne de communication. En effet, reprenons l'expression du signal reçu en amont du corrélateur dans le cas de la figure 5.8. Pour faciliter la formalisation de l'IES, nous nous placerons dans le contexte simplifié suivant : le bruit introduit par la perturbation de canal $p(n)$ est nul de sorte que le seul bruit de la chaîne est le signal audio $x(n)$; les filtres de mise en forme perceptuelle $H(f)$ et blanchissant $\hat{G}(f)$ sont invariants dans le temps [1]. Le signal reçu $\hat{v}(n)$ en amont du corrélateur est le signal audio tatoué $y(n)$ filtré passe-bas, somme du signal audio $x(n)$ et du signal de tatouage $t(n)$ blanchi par le filtre $\hat{G}(f)$:

$$\hat{v}(n) = \hat{g}(n) \star y(n) = \hat{g}(n) \star t(n) + \hat{x}(n), \tag{7.4}$$

où $\hat{g}(n)$ désigne la réponse impulsionnelle du filtre $\hat{G}(f)$ combiné au filtre passe-bas et $\hat{x}(n)$ désigne le signal audio filtré passe-bas puis blanchi par $\hat{G}(f)$, seul bruit de canal. Le signal de tatouage est lui issu du filtrage du signal modulé $v(n) = \sum_{l=0}^{L-1} d_{s_l}(n - lN_s)$ porteur du message

1. Rappelons qu'ils doivent en principe être actualisés sur chaque fenêtre d'analyse de taille N_H et N_G où le signal audio est supposé stationnaire.

numérique $\{s_l\}_{l\in[\![0,L-1]\!]}$ par $H(f)$ de sorte que :

$$\hat{v}(n) = \alpha\hat{g}(n) \star h(n) \star \left(\sum_{l=0}^{L-1} d_{s_l}(n - lN_s)\right) + \hat{x}(n). \tag{7.5}$$

Lors de la détection du symbole s_{L_0}, le signal reçu est observé sur le L_0-ième temps symbole et peut être écrit, par linéarité du filtrage, sous la forme :

$$\hat{v}(n) = \underbrace{\alpha\hat{g}(n) \star h(n) \star d_{s_{L_0}}(n - L_0N_s)}_{w_{L_0}(n)} + \underbrace{\alpha \sum_{l\neq L_0} \hat{g}(n) \star h(n) \star d_{s_l}(n - lN_s)}_{i(n)} + \hat{x}(n), \tag{7.6}$$

où $w_{L_0}(n)$ est le signal porteur du symbole s_{L_0} à l'entrée du corrélateur, $i(n)$ est un bruit additif provenant de l'IES et $\hat{x}(n)$ est toujours le bruit de canal. Ce bruit dépend du temps symbole N_s donc du débit de transmission et du facteur de gain α.

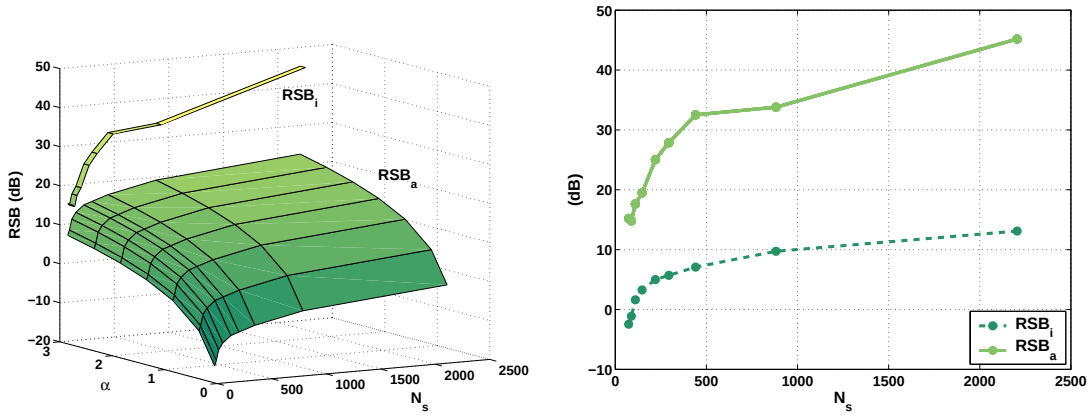
Dans une chaîne de transmission "classique" aux conditions favorables (c'est à dire pour lesquelles le rapport signal à bruit entre le signal utile $w_{L_0}(n)$ et le bruit de canal $\hat{x}(n)$ est élevé), le bruit introduit par l'IES, s'il est plus puissant que le bruit $\hat{x}(n)$, peut nettement dégrader les performances de la transmission. Dans le cas contraire, l'IES pourra être négligée devant le bruit audio $\hat{x}(n)$. Il convient donc d'évaluer les performances de la chaîne de tatouage en terme de RSB.

Deux RSB caractérisent les conditions de transmission proposées par le système de tatouage : d'une part, le rapport de puissance entre le signal porteur de l'information $w_{L_0}(n)$ et le bruit audio blanchi $\hat{x}(n)$ [2] et le rapport de puissance entre ce même signal $w_{L_0}(n)$ et le bruit provenant de l'IES $i(n)$. Le premier sera noté RSB$_a$, le second RSB$_i$. Ces RSB doivent être calculés dans l'***espace de détection*** $\mathfrak{R}_{détection}$, défini comme l'espace engendré par les vecteurs du dictionnaire de réception. En effet, la détection exploitant la projection du signal reçu $\hat{v}(n)$ sur les vecteurs du dictionnaire, toute composante du bruit audio $\hat{x}(n)$ ou de l'IES $i(n)$ en dehors de cet espace n'est pas significative (puisqu'elle sera éliminée par le corrélateur). La mesure des RSB dans cet espace constitue donc une statistique suffisante. Ces RSB dépendent du facteur de gain α [3], des performances des filtres $H(f)$ et $\hat{G}(f)$, du temps symbole N_s (qui modifie l'énergie consacrée à la transmission de chaque symbole) et de la bande passante F_c (désormais fixée à $5,5$ kHz). Nous avons donc mesuré les RSB de la chaîne de tatouage dans le contexte de simulation suivant : les 20 signaux de la base de test ont été tatoués par une séquence de 5000 bits à l'aide d'un dictionnaire de deux vecteurs antipodaux. La figure 7.2 (a) présente les valeurs moyennes du RSB$_a$ et du RSB$_i$ obtenues en fonction du facteur de gain α et du temps symbole N_s. Ce temps symbole coïncide pour l'occasion avec le temps bit N_b ; sa plage de variation a été choisie en faisant varier le débit de transmission $R = \dfrac{F_e}{N_b}$ de 100 à 600 bit/s. La figure 7.2 (b) met l'accent sur les valeurs des RSB lorsque le gain α est fixé à $1,5$ (valeur permettant d'obtenir en moyenne un tatouage perceptible mais non gênant d'ODG -1).

La figure 7.2 (a) met en évidence l'échelle de valeurs prises par les RSB : le rapport entre l'information de tatouage et le bruit audio varie d'environ -20 à 25 dB, celui entre l'information

2. Ce RSB intervient dans le calcul de la probabilité d'erreur de transmission donnée 4.4. Il ne doit pas être confondu avec le RMS (rapport masque à signal) donnant le rapport de puissance entre le signal de tatouage $t(n)$ et le signal audio $\hat{x}(n)$.

3. En réalité seul le RSB$_a$ entre le tatouage et l'audio dépend du gain. En effet, le RSB$_i$ est indépendant de α puisque $w_{L_0}(n)$ et $i(n)$ sont tous deux proportionnels à α.



(a) (b)

Figure 7.2 – Évaluation des rapports signaux à bruit dans la chaîne de tatouage, le RSB_a provenant du signal audio et le RSB_i issu de l'interférence entre symboles : (a) Évolution du RSB_a et du RSB_i en fonction du gain α et du temps symbole N_s ; le RSB_i étant indépendant de α n'est représenté que pour $\alpha = 3$, (b) variations du RSB_a et RSB_i en fonction de N_s pour un gain fixé à $1,5$.

de tatouage et l'IES entre 15 et 45 dB. Les RSB diminuent lorsque le temps symbole diminue, autrement dit lorsque le débit augmente. Pour autant, quels que soient les paramètres de fonctionnement de la chaîne (gain, temps symbole et débit), le RSB lié au bruit audio est toujours inférieur à celui lié à l'IES, ce qui est flagrant figure 7.2 (b) lorsque le gain est fixé à $1,5$. Nous pourrons donc supposer que l'IES dans la chaîne de tatouage est négligeable devant le bruit introduit par le signal audio.

Pour conclure, l'orthogonalité et l'égalité des puissances des vecteurs du dictionnaire de réception étant obtenues pour de faibles valeurs du temps symbole et l'IES pouvant être supposée négligeable, nous pourrons par la suite modéliser le système de tatouage par une chaîne de communication BBAG au modèle adapté.

7.2 Performances expérimentales

Cette section fait état des performances expérimentales du système de tatouage de référence obtenu par tatouage du corpus de 20 signaux selon le protocole expérimental établi précédemment. Plusieurs objectifs sont proposés dans cette étude. Le premier est de confirmer l'efficacité du système de tatouage et plus particulièrement du module de contrôle adaptatif du gain, en terme d'inaudibilité. Le système sera donc simulé avec et sans utilisation du module de contrôle. Le deuxième est d'établir l'efficacité du récepteur par filtre blanchissant en le comparant aux performances de l'égaliseur de Larbi, présenté section 3.2.1. Le dernier est d'établir les performances des modulations M-aire et CDMA dans un contexte de tatouage sur des signaux réels. L'étude des performances théoriques de la chaîne de tatouage a d'ores et déjà établi l'influence du nombre de vecteurs M du dictionnaire sur les performances de la chaîne de tatouage. Dans le cas des modulations M-aire, augmenter M accroît la fiabilité de transmission ; dans le cas des modulations CDMA, la probabilité d'erreur est indépendante de M. Trois cas

ODG	M-aire/CDMA K = 1	M-aire K = 4, M = 16	CDMA K = 4, M = 16
Sans module de contrôle	$-0,95$ $(0,62)$	$-0,94$ $(0,77)$	$-0,94$ $(0,76)$
Avec module de contrôle	$-1,0$ $(0,07)$	$-1,0$ $(0,07)$	$-1,0$ $(0,07)$

Table 7.1 – Valeurs moyennes de l'ODG et écart-type (entre parenthèses) pour le système de tatouage en fonction de la modulation et de l'utilisation du module de contrôle de l'inaudibilité.

de figures seront donc considérés : le cas d'une modulation binaire antipodale (commune aux modulations M-aire et CDMA lorsque le dictionnaire est constitué de $M = 2$ vecteurs antipodaux), le cas d'une modulation M-aire pour un dictionnaire biorthogonal constitué de $M = 16$ vecteurs ($K = 4$) et le cas d'une modulation CDMA pour un dictionnaire orthogonal de $M = 4$ ($K = 4$) vecteurs. Ces trois cas de figure sont respectivement référencés par : *M-aire/CDMA K=1, M=2*, *M-aire K=4, M=16* et *CDMA K=4, M=16*.

Les quatre critères de performance, à savoir l'inaudibilité, le TEB dans le cas d'un canal sans perturbations, la robustesse du système aux perturbations et le coût en temps de calcul, sont analysés pour les modulations considérées et les deux récepteurs.

7.2.1 Inaudibilité

L'inaudibilité est évaluée grâce à la mesure de l'ODG par l'algorithme PEAQ. Nous avons paramétré le système de tatouage de sorte à garantir un tatouage inaudible mais non gênant, caractérisé par une ODG de -1.

L'objectif de cette étude est de mettre en évidence l'efficacité du module de contrôle adaptatif d'inaudibilité. Plusieurs configurations du système sont donc envisagées. Le système est simulé dans un premier temps sans le module de contrôle. Le facteur d'amplitude α est alors fixé à $1,5$, cette valeur permettant d'obtenir une ODG moyenne de -1. Il est ensuite simulé avec le module de contrôle paramétré pour obtenir localement une ODG de -1. La table 7.1 présente les résultats obtenus pour les 3 modulations envisagées [4] : ces résultats sont les valeurs moyennes des ODGs obtenues par tatouage du corpus des 20 signaux par une séquence de 100000 bits à 250 bit/s et leurs écarts-types présentés entre parenthèses.

Ces résultats révèlent la performance du système de tatouage avec contrôle d'inaudibilité en terme de transparence du tatouage, lorsque celle-ci est mesurée par l'algorithme PEAQ : lorsque le module de contrôle n'est pas utilisé, la paramétrisation du système permet d'obtenir un tatouage inaudible mais non gênant (puisque α a été choisi dans cet objectif). L'écart type des ODGs de chaque signal du corpus rappelle néanmoins la très forte dépendance du système au signal audio qui avait conduit à l'implémentation du module de contrôle adaptatif d'inaudibilité. Lorsque ce module est actif, l'ODG atteint là encore la valeur escomptée -1, mais permet surtout une très nette diminution de l'écart-type. Le facteur d'amplitude α est effectivement adapté au signal audio pour faire tendre la construction du tatouage vers un signal inaudible mais non gênant.

Des tests d'écoute informels ont également été réalisés. Pour la plupart des signaux, la transparence du tatouage est jugée très acceptable ; néanmoins nous avons constaté qu'un des signaux audio (le signal *"preacher"*) présente quelques distorsions auditives locales. Ces distorsions sont

4. L'inaudibilité ne dépend pas ici du récepteur choisi.

sans doute dues aux imperfections de la fenêtre d'oubli utilisée pour atténuer les variations brutales du gain et que nous avons déjà évoquées section 6.4.3. Une analyse plus détaillée des défauts du module de contrôle d'inaudibilité pour ce signal devra donc être effectuée, que nous n'avons malheureusement pas encore le temps d'effectuer. Par la suite, les tests d'écoute réalisés sur tous les systèmes de tatouage révèleront les mêmes conclusions.

7.2.2 TEB pour un canal sans perturbations

Cette section s'intéresse aux TEBs du système de tatouage dans le cas d'un canal ne présentant pas de perturbations ; elle propose de mettre en évidence l'apport des modulations M-aire et CDMA sur les performances du système, de comparer l'efficacité du récepteur par filtrage blanchissant par rapport à l'égaliseur proposé par Larbi mais également d'étudier l'influence du module de contrôle d'inaudibilité sur les performances du système. En effet, ce dernier, en adaptant le facteur d'amplitude α pour garantir l'inaudibilité locale modifie les conditions de transmission du système de tatouage et donc influe sur la fiabilité de transmission.

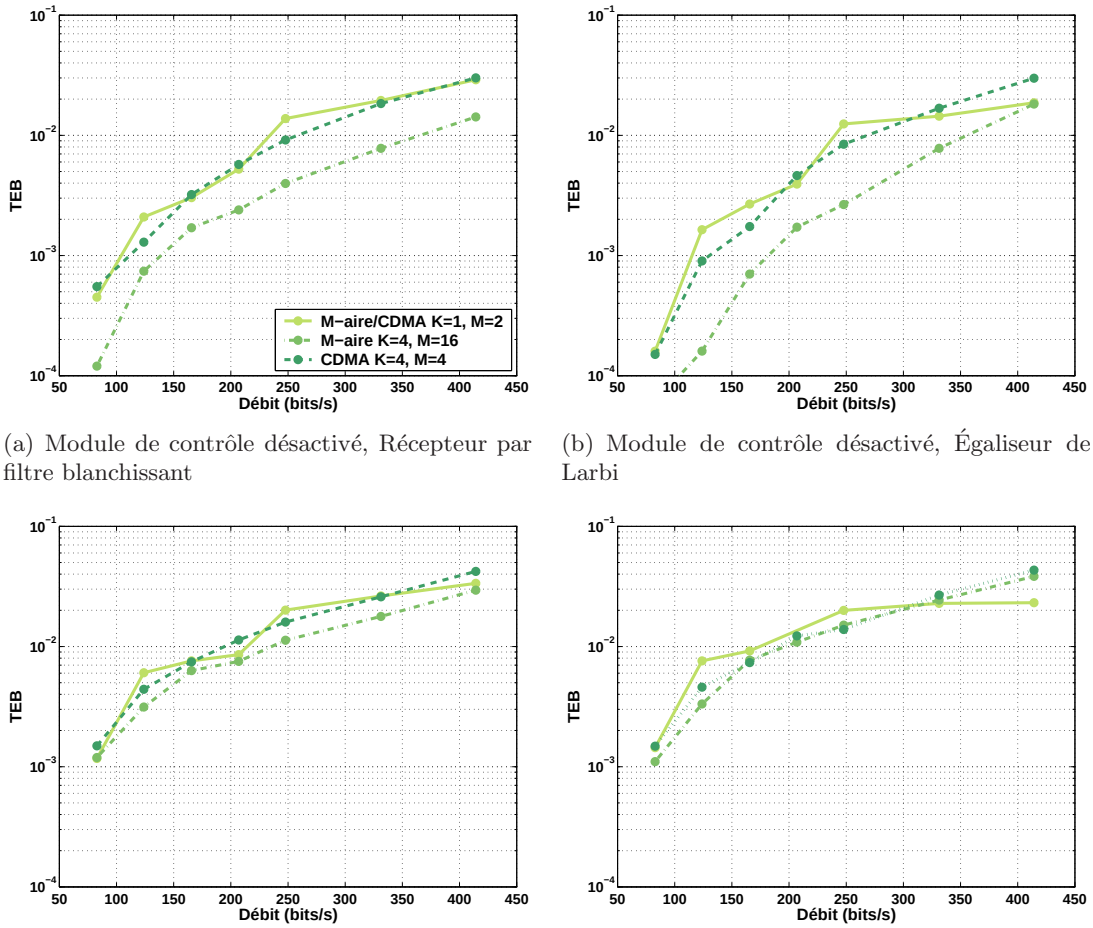
Le système de tatouage a donc été simulé sur le corpus de 20 signaux lorsque le module de contrôle d'inaudibilité est ou non activé, pour les trois configurations de modulation envisagées précédemment et pour les deux récepteurs. Les TEBs obtenus en fonction du débit de transmission $R = \dfrac{KF_e}{N_s}$ pour ces différentes configurations sont présentés figure 7.3. Les figures (a) et (b) s'intéressent au cas d'un module de contrôle d'inaudibilité désactivé tandis que les figures (c) et (d) mettent en œuvre ce module de contrôle. Les figures (a) et (c) sont obtenues pour un récepteur par filtre blanchissant et dictionnaire de réception filtré, tandis que les figures (b) et (d) font le pendant dans le cas de l'égaliseur de Larbi. Pour chaque figure, les trois configurations de modulation sont considérées.

Efficacité du récepteur par filtre blanchissant

Lorsque le module de contrôle d'inaudibilité est désactivé, les courbes montrent que les performances du récepteur par filtre blanchissant sont globalement (quelle que soit la modulation) inférieures à celle de l'égaliseur de Larbi : les TEBs avec filtrage blanchissant sont environ $1,5$ fois supérieurs à ceux de l'égaliseur. Lorsque le module de contrôle d'inaudibilité est activé, les différences de TEBs s'atténuent ; les deux récepteurs sont alors d'égale performance.

Influence du module d'inaudibilité

L'introduction du module de contrôle induit une nette augmentation des TEBs et ce, quel que soit le récepteur ou la modulation. En effet, le corpus contient plusieurs signaux très sensibles à l'introduction du tatouage, pour lesquels le gain α doit être fortement décru (par rapport à la valeur $1,5$ utilisée pour un réglage moyen) pour garantir une ODG de -1. Pour ces signaux, les TEBs obtenus pour la valeur adaptée de α sont donc nettement supérieurs à ceux obtenus avec le système sans contrôle d'inaudibilité. A titre d'exemple, le signal test (a) étudié section 6.4.3 obtient à 206 bit/s (lorsque le récepteur est le filtre blanchissant) un TEB de $8,8.10^{-3}$ lorsque le module de contrôle est actif contre $0,4.10^{-3}$ lorsqu'il est inactif, soit une augmentation de 20%. Ces signaux influencent fortement le TEB moyen et provoquent l'augmentation de celui-ci. Cette augmentation est d'autant plus nette pour le récepteur basé sur l'égaliseur de Wiener.



(a) Module de contrôle désactivé, Récepteur par filtre blanchissant

(b) Module de contrôle désactivé, Égaliseur de Larbi

(c) Module de contrôle actif, Récepteur par filtre blanchissant

(d) Module de contrôle actif, Égaliseur de Larbi

Figure 7.3 – TEBs du système de tatouage non informé de référence en fonction du débit de transmission dans le cas d'un canal sans perturbation pour plusieurs configurations de modulation, de contrôle d'inaudibilité, de dictionnaire et de récepteurs. Les modulations sont la modulation M-aire ou CDMA ; le dictionnaire contient soit 2 vecteurs antipodaux ($K = 1, M = 2$), 16 vecteurs biorthogonaux ($K = 4, M = 16$), soit 4 vecteurs orthogonaux ($K = 4, M = 4$). Le récepteur est basé soit basé sur le filtrage blanchissant, soit sur l'égaliseur de Larbi.

Apport des modulations

Lorsque le module de contrôle d'inaudibilité est inactif, l'apport des modulations sur les performances du système de tatouage est conforme aux courbes théoriques établies section 4.4 : les TEBs obtenus par les modulations CDMA sont indépendants du nombre de vecteurs du dictionnaire, et ce quel que soit le récepteur ; les TEBs peuvent par contre être décrus en exploitant des modulations M-aire et en jouant sur le nombre de vecteurs du dictionnaire. Pour ces dernières, la figure (a) montre que choisir un encodage de $K = 4$ permet de diviser les TEBs d'environ 2 par rapport à un encodage de $K = 1$.

Lorsque le module de contrôle est activé, ces conclusions sont moins nettes : les TEBs quels que soient la modulation et le récepteur choisis ont tous un ordre de grandeur relativement simi-

laire. Les modulations M-aire pour $M = 16$ restent globalement les plus performantes des trois configurations envisagées mais le gain apporté sur les TEBs en comparaison d'une configuration sans encodage est inférieur à 2. Le module de contrôle d'inaudibilité a donc pour effet d'atténuer les écarts entre les performances des différentes modulations. Cette atténuation est pourtant prévisible : en choisissant de garantir une ODG de -1, le facteur d'amplitude α adaptatif obtenu pour de nombreux signaux audio du corpus devient inférieur à la valeur $1,5$ utilisée dans le système de tatouage sans application du module de contrôle d'inaudibilité ; l'activation du module de contrôle dégrade donc les conditions de transmission en terme de RSB. Conformément aux performances théoriques de la chaîne exposées figure 4.4 (a), plus ce RSB est petit, plus l'apport de l'augmentation de la taille du dictionnaire sur la probabilité d'erreur est faible. Malgré cette atténuation de l'écart de performances entre les modulations, nous conserverons par la suite la configuration *M-aire* $K = 4$, $M = 16$ (plus performante que la configuration *M-aire/CDMA*, $K = 1$, $M = 2$) et la configuration *CDMA* $K = 4$, $M = 4$. Ces deux configurations serviront de référence tout au long de ce document, pour évaluer l'apport des stratégies d'insertion informées développées par la suite sur les performances du système.

7.2.3 Robustesse de la chaîne de transmission

Les figures 7.4 concernent la robustesse du système lorsque le débit de transmission est fixé à 206 bit/s [5] : elles cherchent à établir l'efficacité du système de tatouage pour les différentes modulations considérées (modulation M-aire et CDMA pour un encodage $K = 4$) et l'impact du module de contrôle d'inaudibilité sur les performances. Les perturbations considérées et établies par le protocole expérimental sont rappelées et numérotées table 7.2.

Numéro	Perturbations	Numéro	Perturbations
1	Sans perturbations	**7**	Sous échantillonnage (22050 kHz)
2	Compression MPEG 96 kbit/s	**8**	Quantification (12 bits)
3	Compression MPEG 64 kbit/s	**9**	Ajout d'un bruit blanc (30 dB)
4	Filtrage passe-haut* (250 Hz)	**10**	Amplification (réduction de 50%)
5	Filtrage passe-bas* (9 kHz)	**11**	Compression de dynamique*
6	Ajout d'écho		

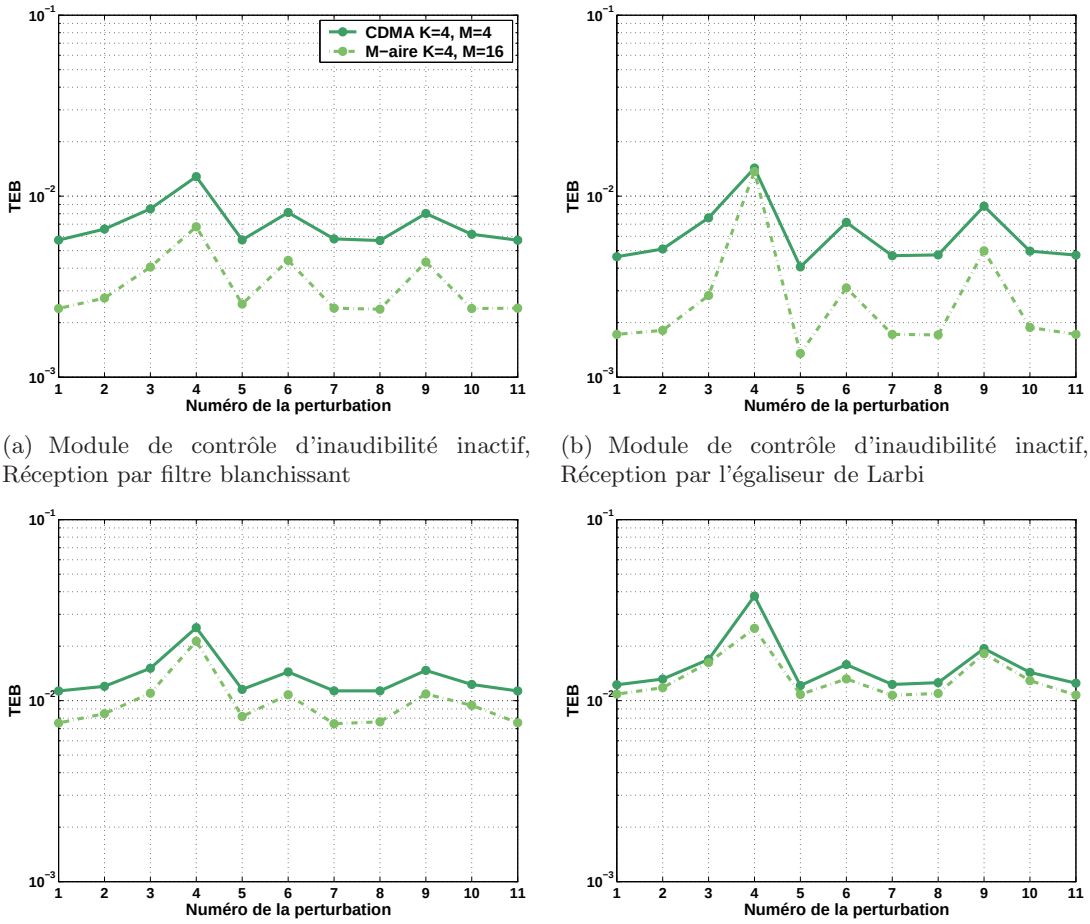
Table 7.2 – Liste des perturbations et de leur numéro associé. Les perturbations indiquées par une * sont réalisées par le Stirmark Audio.

La figure 7.4 (a) et (b) présente les TEBs des deux modulations considérées en fonction du numéro de la perturbation lorsque le module d'inaudibilité est inactif. La figure (a) s'attache au récepteur par filtre blanchissant et la figure (b) à l'égaliseur de Larbi. La figure 7.4 (c) et (d) fait le pendant des figures (a) et (b) dans le cas où le module de contrôle d'inaudibilité est activé.

Ces courbes font état d'une robustesse très acceptable du système aux perturbations et ce quelles que soient la modulation ou l'activité du module de contrôle. Les TEBs obtenus pour chaque configuration sont relativement proches des valeurs sans perturbations. Deux groupes de perturbations peuvent être distingués :

- les perturbations peu dégradantes tels que les filtrages passe-bas (numéro 5), le sous-échantillonnage (7), la quantification (8), l'amplification (10) et les compressions de dy-

5. Cette valeur de débit est choisie au regard de l'ensemble des résultats de la thèse pour correspondre à des TEBs de l'ordre de 5.10^{-3}. En effet, la fiabilité des résultats étant de l'ordre de 10^{-3}, les mesures pour des TEBs inférieures à 10^{-3} ne sont pas significatives.



(a) Module de contrôle d'inaudibilité inactif, Réception par filtre blanchissant

(b) Module de contrôle d'inaudibilité inactif, Réception par l'égaliseur de Larbi

(c) Module de contrôle d'inaudibilité actif, Réception par filtre blanchissant

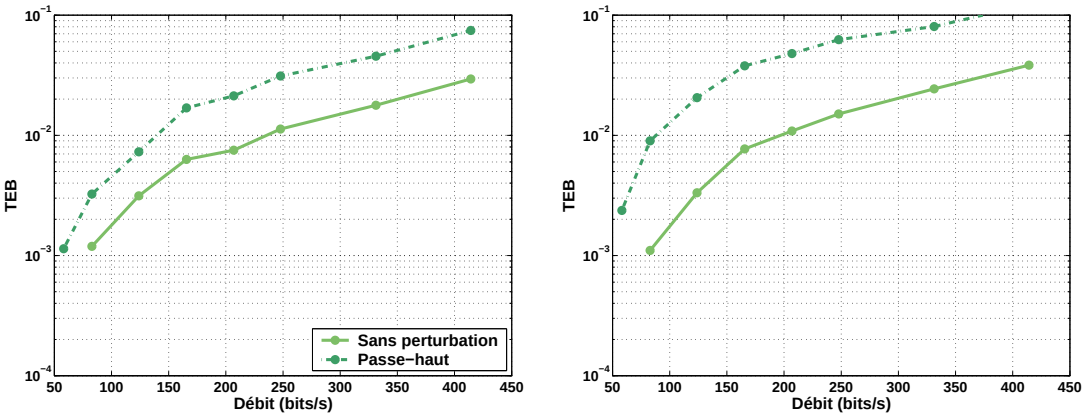
(d) Module de contrôle d'inaudibilité actif, Réception par l'égaliseur de Larbi

Figure 7.4 – Évolution du TEB à 206 bit/s en fonction des perturbations (classées par ordre croissant) pour différentes modulations (M-aire et CDMA) et deux récepteurs : (a) le filtre blanchissant et (b) l'égaliseur de Wiener.

namique (11),

- les perturbations légèrement dégradantes que sont la compression MPEG à 64 kbit/s (3), le filtrage passe-haut (4), l'ajout d'écho (6) et l'ajout de bruit (9).

Lorsque le module de contrôle d'inaudibilité est activé, la perturbation la plus dégradante est, quel que soit le récepteur ou la modulation, le filtrage passe-haut. Elle multiplie par environ $2,8$ le TEB par rapport au cas sans perturbations. La figure 7.5 complète donc cette analyse de la robustesse du système aux perturbations ; elle compare, pour la modulation M-aire $M = 16$, $K = 4$ conservée par la suite, les TEBs en fonction du débit dans le cas d'un canal sans perturbation et d'un canal avec filtrage passe-haut (c'est à dire la perturbation la plus dégradante). Les courbes sont proposées pour les deux types de récepteur lorsque le module de contrôle d'inaudibilité est activé. Au regard de l'augmentation des TEBs introduite par le filtrage passe-haut, une transmission de TEB 10^{-3} et robuste aux perturbations peut donc être obtenue pour un débit de l'ordre de 50 bit/s.



(a) Module de contrôle d'inaudibilité actif, Réception par filtre blanchissant

(b) Module de contrôle d'inaudibilité actif, Réception par l'égaliseur de Larbi

Figure 7.5 – Comportement des TEBs en fonction du débit face à un canal dégradé par un filtrage passe-haut, comparé à un canal sans perturbations, pour les deux récepteurs : (a) le filtre blanchissant et (b) l'égaliseur de Wiener. La modulation est M-aire avec $M = 16$, $K = 4$.

Émetteur	Sans module de contrôle	Avec module de contrôle
Ratio	23%	78 %

Table 7.3 – Coût en temps de calcul par rapport au temps réel du module de contrôle d'inaudibilité à l'émetteur.

7.2.4 Coût de calcul

Le coût du système en temps de calcul pouvant être un facteur de choix déterminant pour l'application temps-réel du système, nous proposons dans cette section une estimation de ce coût. Cette estimation est le rapport entre le temps nécessaire au traitement du signal et la durée du signal audio tatoué lorsque le système de tatouage est implémenté en C et est simulé sur un PC de bureau dont les caractéristiques sont données dans le protocole expérimental.

Nous souhaitons évaluer dans un premier temps le coût de l'ajout du module de contrôle d'inaudibilité. La table 7.3 fait état des résultats obtenus pour les deux configurations d'émetteur. Elle montre que, puisque l'émetteur non informé est de coût très faible, il n'est que peu complexifié par l'ajout du module de contrôle d'inaudibilité et reste implémentable en temps-réel.

Dans un second temps, nous proposons de comparer le coût du récepteur par filtre blanchissant à celui de l'égaliseur. Ce coût dépend du nombre M de vecteurs dans le dictionnaire. La figure 7.6 présente de fait ses variations en fonction de M pour les deux récepteurs. L'implémentation de l'égaliseur de Larbi apparaît être réalisable en temps-réel et être peu sensible au nombre de vecteurs du dictionnaire. Le coût du récepteur par filtre blanchissant est globalement supérieur à celui de l'égaliseur. De plus, le coût de son implémentation augmente exponentiellement avec M : en effet, contrairement à l'égaliseur, le récepteur par filtre blanchissant nécessite le calcul d'un dictionnaire de réception, résultat du filtrage des M vecteurs du dictionnaire sur chaque temps symbole. Une mise en œuvre temps-réel du récepteur par filtre blanchissant n'est envisageable pour cette configuration de PC que si le dictionnaire contient

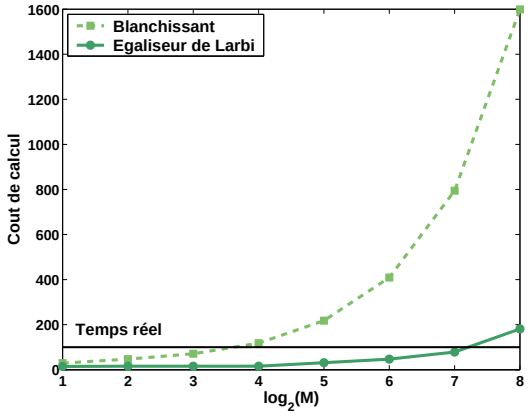


Figure 7.6 – Variations du coût de réception dans le cas du filtre blanchissant et de l'égaliseur de Larbi en fonction du nombre M de vecteurs du dictionnaire.

au plus 16 vecteurs. Sa faisabilité en temps-réel est donc possible à la condition d'utiliser un dictionnaire de faible dimension.

Conclusion

Ce chapitre s'est attaché à la mise en œuvre effective du système de tatouage sur des signaux audio réels. Cette mise en œuvre a permis d'évaluer les conditions de transmission, en terme de rapport signal à bruit et d'interférences entre symbole, de la chaîne de tatouage, justifiant les hypothèses effectuées pour sa conception. Il a surtout permis d'évaluer les performances de la chaîne de tatouage et de les comparer à un système de la littérature (de conception très similaire). Les critères de performances mesurés révèlent une transparence très correcte du tatouage : le tatouage obtenu est audible mais non gênant et ce quel que soit le signal audio considéré. Les TEBs en fonction du débit ont également été évalués : ils montrent qu'une transmission de TEB 10^{-3} peut être obtenue pour un débit de l'ordre de 50 bit/s, et ce quelle que soit la perturbation non désynchronisante considérée. Ce TEB est atteint grâce à l'utilisation de modulations M-aire, dont les performances sont meilleures que les modulations CDMA. Il a été montré que l'introduction de perturbations influence peu les TEBs du système : dans le pire des cas, une augmentation de l'ordre de 46% est à prévoir. Le coût en temps de calcul a également été évalué montrant la faisabilité d'une application temps-réel à condition de choisir un dictionnaire d'émission contenant au plus une dizaine de vecteurs.

Conclusion

Dans cette seconde partie, nous avons proposé et justifier la conception d'un système de tatouage audio non informé sous la seule hypothèse que le signal de tatouage $t(n)$ est ajouté au signal audio $x(n)$.

Le système proposé a été conçu sous la forme d'une chaîne de communication dont le schéma de principe est rappelé figure 7.7. Son principe est le suivant :

- L'émetteur insère une information binaire (représentée soit sous la forme de symbole soit sous la forme d'un K-uplet binaire) dans le signal audio hôte $x(n)$. Cette insertion débute par une étape de modulation basée sur un dictionnaire d'émission $\mathcal{D} = \{\mathbf{d}_m\}_{m\in[\![0,M-1]\!]}$ contenant M vecteurs, choisis blancs étalés dans la bande fréquentielle $[0; F_c]$ (avec $F_c = 5,5$ kHz pour les besoins de robustesse à la compression MPEG), de puissance unité et vérifiant des propriétés d'orthogonalité. Deux modulations par étalement de spectre, adaptées au système de tatouage, ont été envisagées : les modulations M-aire et les modulations CDMA. Le signal modulé $v(n)$ est ensuite mis en forme spectralement à l'aide d'un filtre de mise en forme perceptuelle $H(f)$ et d'un gain α, pour obtenir le signal de tatouage $t(n)$. Ce filtre est conçu de sorte que la densité spectrale de puissance de $t(n)$ coïncide avec un seuil de masquage, obtenu après analyse du signal audio par un Modèle PsychoAcoustique (MPA). Ce seuil est une limite fréquentielle caractérisant la contrainte d'inaudibilité du signal $t(n)$ en présence de $x(n)$. Le facteur de gain α permet lui de contrôler la distorsion perceptuelle introduite par le tatouage via un module de contrôle d'inaudibilité (qui n'est pas représenté sur la figure). Ce module permet de choisir le gain α spécifique au signal audio en cours de traitement permettant d'atteindre un degré de transparence donné. Pour ce faire, il exploite un algorithme de mesure objective de la différence perceptuelle entre deux signaux qui rétroagit sur le gain α. Le signal audio tatoué $y(n)$ est finalement obtenu en ajoutant le signal audio $x(n)$ et le signal de tatouage $t(n)$.

- Le canal est le siège des perturbations que peut classiquement subir un signal audio et qui ont été évoquées précédemment. Dans cette partie, nous nous sommes limités au cas des perturbations non désynchronisantes, qui feront l'objet du chapitre 11. Ce canal conduit au signal audio tatoué détérioré $\hat{y}(n)$.

- Le récepteur permet l'extraction de l'information émise étant donné le signal audio tatoué dégradé $\hat{y}(n)$, reçu en sortie de canal. Il pré-traite le signal reçu par un filtre Passe-Bas (PB) de fréquence de coupure F_c puis utilise un filtre blanchissant $\hat{G}(f)$ calculé par un algorithme de Prédiction Linéaire (PL). Le but de ce filtre est de blanchir le bruit prédominant de la chaîne, à savoir le signal audio $x(n)$ lui-même. La décision est ensuite basée sur le calcul de la corrélation entre le signal reçu filtré $\hat{v}(n)$ et un dictionnaire de réception $\hat{\mathcal{D}}$, contenant l'ensemble des vecteurs du dictionnaire d'émission filtrés par les deux filtres de la chaîne : le

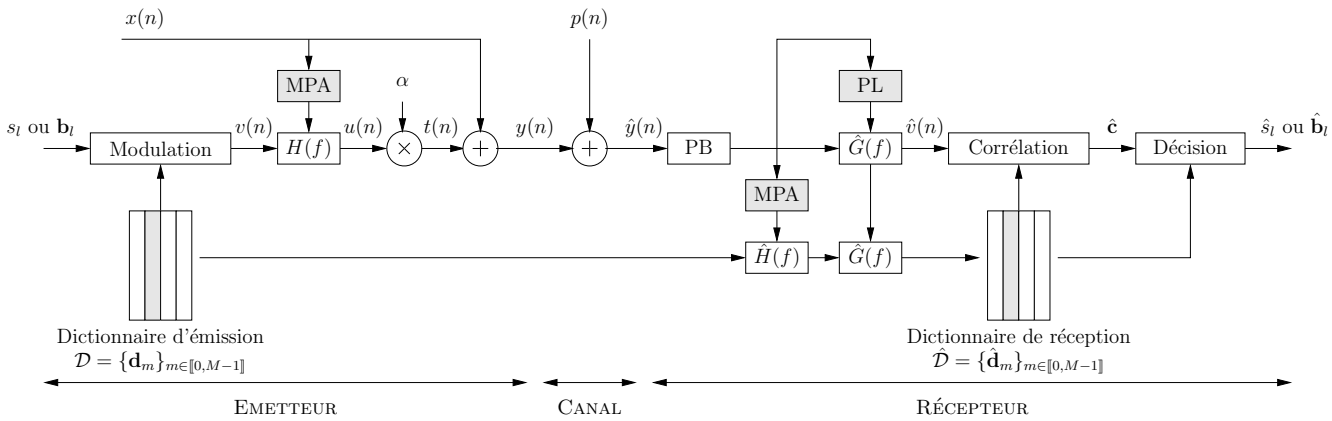


FIGURE 7.7 – Schéma du système de tatouage. Le module de contrôle d'inaudibilité choisissant α en fonction de $x(n)$ et $y(n)$ a été omis pour des raisons de lisibilité.

filtre $\hat{H}(f)$ (estimation de $H(f)$ calculé à partir du signal audio reçu $\hat{y}(n)$) et le filtre $\hat{G}(f)$. Cette procédure permet d'extraire l'information reçue en utilisant deux détecteurs sous-optimaux spécifiques aux modulations choisies : un détecteur par corrélation dans le cas des modulations M-aire et un démodulateur par décorrélation dans le cas des modulations CDMA. L'information reçue est alors traduite par le détecteur soit sous la forme d'un symbole $\hat{s}_l$ soit sous la forme d'un vecteur $\hat{\mathbf{b}}_l$.

La paramétrisation de ce système a fait l'objet d'une étude détaillée pour choisir les valeurs de paramètres les plus adaptées à l'optimisation des performances de ce système. Les performances effectives ont ensuite été évaluées sur un corpus de 20 signaux selon les quatre critères du protocole expérimental établi dans la première partie. Ces critères montrent une transparence très acceptable du tatouage (qualifiée d'audible mais non gênante). Ils proposent une évaluation du TEB en fonction du débit en l'absence de perturbations de canal, servant de référence comparative pour les systèmes de tatouage développés par la suite. La robustesse de notre système aux perturbations non désynchronisantes se révèle être très correcte. Une transmission de TEB 10^{-3} peut de fait être obtenue à un débit d'environ 50 bit/s, quelle que soit la perturbation non désynchronisante considérée. Les TEBs de notre système ont également été comparés à un système de l'état de l'art, dont le schéma de principe est très similaires à notre système. Il diffère essentiellement par l'étape de réception basée non plus sur un filtre blanchissant mais sur une procédure d'égalisation : l'étude révèle des performances très similaires au système de l'état de l'art. Pour finir, une implémentation en C de notre système a permis de montrer sa faisabilité en temps-réel.

Les performances de ce système peuvent être significativement améliorées à la condition de prendre en compte la connaissance *a priori* du bruit audio à l'émetteur et de faire tendre la stratégie d'insertion du tatouage vers une stratégie informée. Tel est l'objectif de la troisième partie de ce manuscrit.

Troisième partie

Vers un système de tatouage informé

Introduction

Exploiter la connaissance *a priori* du signal audio à l'émetteur pour concevoir un système de tatouage performant est l'enjeu majeur des recherches actuelles dans le domaine du tatouage et l'objectif de cette thèse : le signal audio est mis à profit pour établir une stratégie d'insertion adaptée, permettant de construire un tatouage inaudible et robuste aux perturbations de canal.

Différentes stratégies d'insertion ont d'ores et déjà été proposées dans la littérature : elles s'attachent aussi bien à des modèles théoriques de la chaîne de tatouage (sous la forme d'une chaîne de communication avec information adjacente similaire à celle étudiée par Costa [Cos83]) qu'à des mises en œuvre pratiques sur des signaux réels. Bien que leurs efficacités en terme de TEB et de robustesse aux perturbations soient acquises, elles montrent de nombreuses faiblesses dans la prise en compte de la contrainte d'inaudibilité. Pour certaines, l'inaudibilité est volontairement mise de côté pour satisfaire à une robustesse aux perturbations constante ; pour d'autres, l'inaudibilité est contrôlée par un critère d'erreur quadratique moyenne entre le signal audio tatoué et le signal original, que l'on sait ne pas être pertinent. Notre objectif est donc de concevoir un système de tatouage audio qui garantisse en premier lieu l'inaudibilité et s'attache ensuite à offrir une robustesse maximale aux perturbations. Ces deux contraintes doivent être formalisées avec rigueur.

La transparence du tatouage est contrôlée par l'utilisation d'un filtre de mise en forme perceptuelle, issu d'un modèle psychoacoustique, attenant à un gain α ; ce gain α est choisi spécifiquement en fonction du signal audio pour garantir une transparence donnée, fixée par le concepteur du système. Ce choix est effectué par un module de contrôle adaptatif basé sur l'algorithme de mesure objective de la distorsion PEAQ [Uni01]. Une seule contrainte, liée à l'implémentation du filtre de mise en forme perceptuelle, reste donc à satisfaire par le tatouage pour être inaudible : le signal modulé dont il est issu doit être de puissance unité.

Les conditions de détection correcte et robuste de l'information dépendent de la configuration du signal audio qui interfère sur le procédé de détection. Pour prendre en compte la connaissance de ce signal audio à l'émission, nous proposons d'introduire à l'émetteur une copie locale du récepteur, conférant pour l'occasion une structure en boucle fermée au système de tatouage. Cette copie locale (dépendante du récepteur choisi) permet d'estimer les signaux interférants dans le procédé de détection, et donc d'établir les conditions de la transmission correcte et robuste. Du formalisme de ces conditions peut être déduit un modèle de la chaîne de tatouage sur lequel nous pourrons proposées des stratégies d'insertion informée et en établir les performances théoriques. L'estimation de ces conditions de transmission et le modèle qui en découle font l'objet du premier chapitre de cette partie.

Les conditions de transmission correcte et robuste étant dépendantes de la modulation choisie, nous distinguerons dans un second chapitre les modulations M-aires et dans un troisième les mo-

dulations CDMA. L'introduction de la copie locale du récepteur à l'émetteur permet, outre la formalisation du problème de tatouage, de rétroagir sur l'étape de modulation au regard des conditions de transmission. Un tatouage de robustesse maximale aux perturbations sous contrainte d'inaudibilité peut finalement être choisi. Il satisfait à différents problèmes d'optimisation (fonction de la modulation et du formalisme associé à la robustesse) définissant différentes stratégies d'insertion informée que nous nous proposons de résoudre. Les performances de ces stratégies seront établies sur le modèle du système de tatouage puis seront appliquées au système de tatouage en boucle fermée sur des signaux réels pour jauger de leurs efficacités. Ce développement sera proposé à l'identique pour les modulations M-aire au chapitre 9 et les modulations CDMA au chapitre 10.

La robustesse aux opérations désynchronisantes pouvant être établie grâce à un mécanisme de synchronisation indépendant de la stratégie d'insertion, seules les perturbations non désynchronisantes sont considérées dans cette partie.

Chapitre 8

Conditions d'une transmission inaudible et robuste

Pour exploiter la connaissance *a priori* du signal audio à l'émetteur, nous proposons d'introduire une copie locale du récepteur à l'émetteur. Cette copie locale permet d'estimer les signaux intervenant dans le procédé de détection et de rétroagir sur l'étape de modulation qui régit la construction du tatouage. L'estimation des signaux permet d'établir les conditions d'une transmission inaudible et robuste de l'information tatouée ; la rétroaction permet de choisir un signal de tatouage adapté, satisfaisant à ces conditions.

Dans cet objectif, le tatouage n'est plus construit à partir du vecteur du dictionnaire associé au symbole à émettre mais à partir d'une combinaison linéaire des vecteurs du dictionnaire. Modifier la méthode de construction du tatouage ne modifie pas pour autant le procédé de réception ni les conditions de détection de l'information cachée.

8.1 Copie locale du récepteur à l'émetteur

La connaissance du signal audio à l'émetteur est prise en compte en introduisant à l'émetteur une copie locale de l'émetteur qui rétroagit sur l'étape de modulation. La copie locale du récepteur à l'émetteur dépend naturellement du récepteur choisi. Deux récepteurs ont jusqu'à maintenant été considérés : le récepteur par filtre blanchissant associé à un démodulateur par corrélation basé sur le calcul d'un dictionnaire de réception que nous avons établi section 5.2.1 et l'égaliseur proposé par S. Larbi exposé section 3.2.1. Ces deux récepteurs étant adaptés au même schéma d'émission, il paraît intéressant de les prendre tout deux en compte dans la conception d'une stratégie d'insertion informée [1]. Ils seront par la suite référés sous l'appellation *récepteur par filtre blanchissant* et *égaliseur*.

La figure 8.1 rappelle dans un schéma unifié le principe du système de tatouage avec les deux récepteurs considérés. Pour des raisons de lisibilité, le module de contrôle d'inaudibilité n'est pas représenté ici mais sera pris en compte par la suite. De ce système unifié sont déduits deux schémas d'émetteur avec copie locale du récepteur, présentés figure 8.2. Ces deux schémas font état du système lors de l'insertion du symbole s_l (ou du K-uplet $\mathbf{b}_l$) au l-ième temps symbole ;

1. d'autant plus que l'on sait la réception de l'égaliseur temps-réel alors que celle du filtre blanchissant ne l'est pas

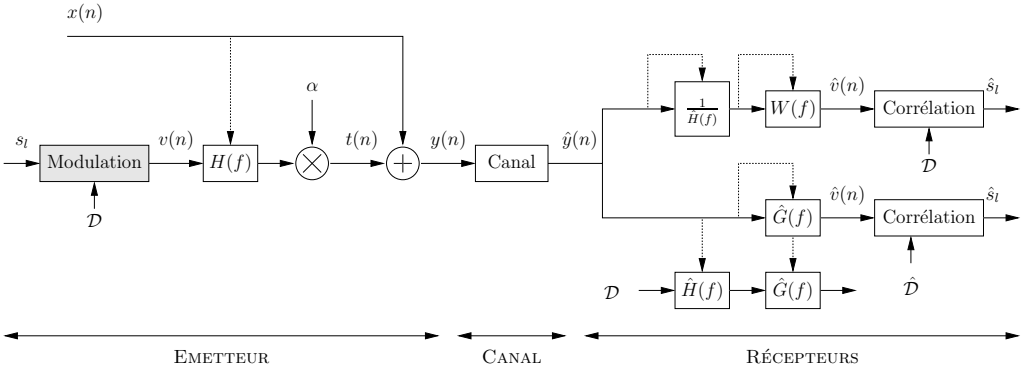


Figure 8.1 – Schéma synthétique du système de tatouage avec les deux récepteurs par filtre blanchissant et égalisation.

par la suite, nous supposerons que ce symbole a pour valeur k (respectivement que le K-uplet a pour valeur $\mathbf{b}$) et nous ne considérerons que l'observation des signaux sur le l-ième temps symbole ; ces observations seront représentées sous forme vectorielle.

La figure 8.2 (a) s'attache au cas du récepteur par filtrage blanchissant : le filtre blanchissant $G(f)$ (théoriquement calculé à partir du signal audio tatoué mais estimé ici à partir du signal audio original) est déplacé en amont de l'additionneur et de l'amplificateur ; ce formalisme met en évidence les signaux intervenant dans le procédé de détection, c'est à dire à l'entrée du démodulateur par corrélation. Il s'agit :

- du signal audio blanchi $\tilde{\mathbf{x}}$,
- du signal modulé $\tilde{\mathbf{v}}$ filtré par le filtre de mise en forme perceptuelle $H(f)$ puis blanchi,
- du signal de tatouage $\tilde{\mathbf{t}} = \alpha\tilde{\mathbf{v}}$ après application des filtres de la chaîne
- et d'une estimation du dictionnaire de réception notée $\tilde{\mathcal{D}}$, résultant du filtrage des formes d'onde du dictionnaire d'émission par $H(f)$ et $G(f)$.

La figure 8.2 (b) présente le schéma de la copie locale du récepteur à l'émetteur lorsque le récepteur est l'égaliseur proposé par Larbi. Là encore, les filtres $\dfrac{1}{\hat{H}(f)}$, inverse du filtre de mise en forme perceptuelle associé au signal audio tatoué $\hat{y}(n)$, et $W(f)$, filtre de Wiener adapté au signal blanchi $\hat{z}(n)$, sont déplacés en amont de l'additionneur et de l'amplificateur. Par construction, $\hat{H}(f)$ est une estimation robuste du filtre de mise en forme perceptuelle $H(f)$ calculé à partir du signal audio original : ainsi $\dfrac{H(f)}{\hat{H}(f)} \approx 1$, ce qui permet de simplifier les étapes de filtrages du signal modulé. Ce schéma permet d'estimer les trois signaux participant à la démodulation par corrélation :

- le signal audio $\tilde{\mathbf{x}}$ blanchi par $\dfrac{1}{H(f)}$ puis filtré par $W(f)$,
- le signal modulé après filtrage de Wiener $\tilde{\mathbf{v}}$
- et le tatouage $\tilde{\mathbf{t}} = \alpha\tilde{\mathbf{v}}$ après traitement par Wiener.

Le corrélateur exploitant le dictionnaire d'émission $\mathcal{D}$[2], l'estimation d'un dictionnaire de

2. puisque le rôle du filtre de Wiener est d'estimer le signal modulé $\mathbf{v}$ au regard du signal reçu $\mathbf{z}$

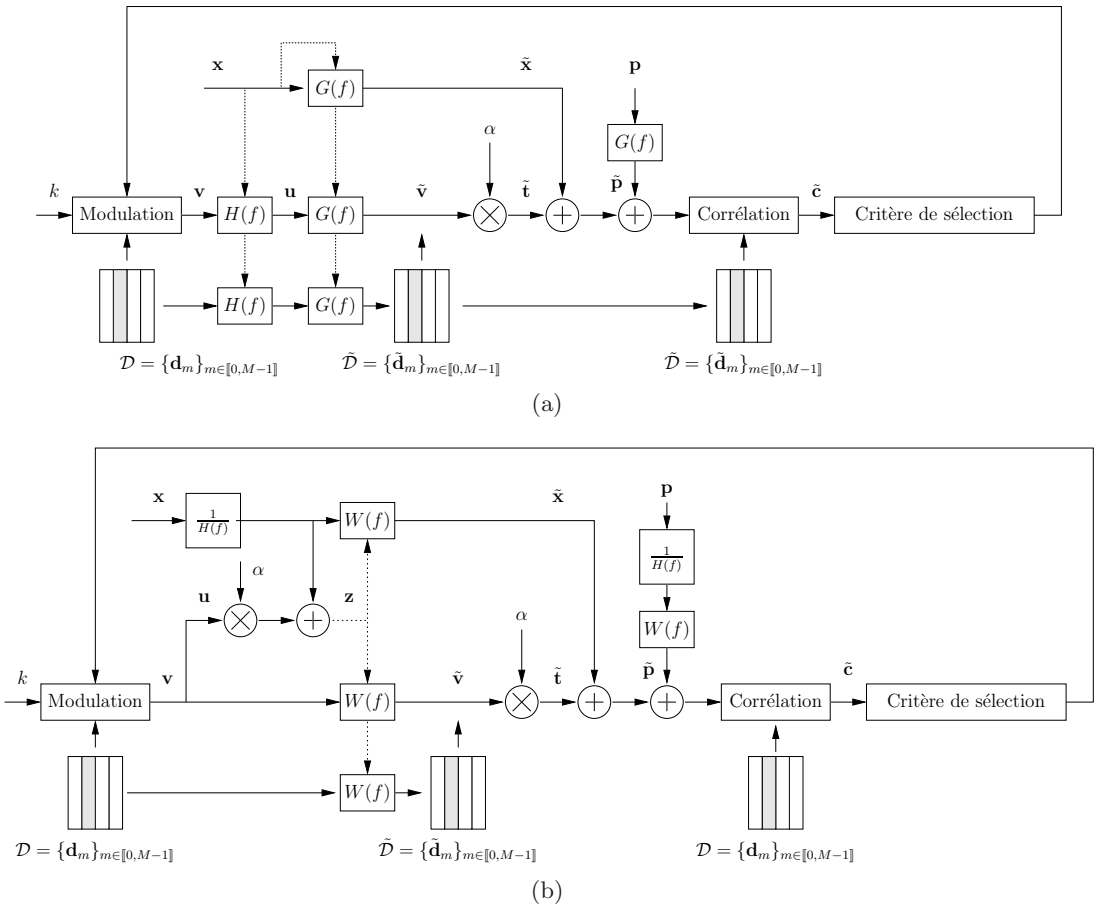


Figure 8.2 – Copie locale du récepteur à l'émetteur dans le cas : (a) du récepteur par filtre blanchissant et dictionnaire filtré, (b) de l'égaliseur.

réception n'est pas nécessaire. Néanmoins, le calcul du dictionnaire $\tilde{\mathcal{D}} = \left\{\tilde{\mathbf{d}}_m\right\}_{m\in[\![0,M-1]\!]}$, contenant l'ensemble des formes d'onde du dictionnaire d'émission filtrées par le filtre de Wiener, nous sera utile.

Pour ces deux schémas, le canal de transmission est pris en compte sous la forme d'un bruit additif ayant subi, aux mêmes titres que le signal audio ou le signal modulé, les traitements du récepteur. Il en est déduit un bruit $\tilde{\mathbf{p}}$ intervenant dans le module de détection.

Un modèle unifié de ces deux stratégies d'insertion informées commence à se dégager. En effet on constate que la configuration des signaux en entrée du corrélateur est très similaire : elle est la somme du signal modulé filtré $\tilde{v}$ construit à partir d'un dictionnaire $\tilde{\mathcal{D}}$, du bruit audio $\tilde{\mathbf{x}}$ et d'un bruit de canal $\tilde{\mathbf{p}}$. Seul diffère le dictionnaire utilisé pour la démodulation par corrélation : $\mathcal{D}$ pour l'égaliseur et $\tilde{\mathcal{D}}$ pour le récepteur par filtre blanchissant. Pour simplifier le formalisme des équations à venir, nous noterons indifféremment ce dictionnaire de réception $\breve{\mathcal{D}} = \left\{\breve{\mathbf{d}}_m\right\}_{m\in[\![0,M-1]\!]}$.

8.2 Conditions d'une transmission inaudible et robuste

L'estimation des signaux intervenant dans le procédé de démodulation par corrélation permet d'établir les conditions à satisfaire par le tatouage $\tilde{\mathbf{t}}$ (et donc par le signal modulé filtré $\tilde{\mathbf{v}} = \frac{1}{\alpha}\tilde{\mathbf{t}}$) pour garantir la transmission inaudible, correcte et robuste de l'information symbolique k ou binaire $\mathbf{b}$. Ces conditions dépendent naturellement des pré-traitements du récepteur (filtrage blanchissant ou filtrage de Wiener) mais surtout de la modulation choisie : nous considérerons donc le cas des modulations M-aire et celui des modulations CDMA. Le facteur de gain α provenant du module de contrôle de l'inaudibilité est supposé connu et constant pendant l'insertion de l'information k [3].

8.2.1 Contrainte d'inaudibilité

L'inaudibilité est garantie par l'utilisation du filtre de mise en forme perceptuelle $H(f)$, issu de l'analyse psychoacoustique du signal audio $\mathbf{x}$, et du choix adéquat du gain α. Elle ne dépend ni du récepteur ni de la modulation choisis.

Le choix du signal modulé $\mathbf{v}$ est limité par l'implémentation du filtre $H(f)$ qui impose un signal de puissance unité. La contrainte d'inaudibilité se traduit alors par la contrainte de puissance suivante que nous avions déjà établie section 5.1.3 :

$$\sigma_v^2 = \frac{1}{N_s}\mathbf{v}^t\mathbf{v} \leq 1. \tag{8.1}$$

8.2.2 Conditions de transmission correcte et robuste

Les conditions d'une transmission correcte et robuste dépendent du récepteur et de la modulation choisis. Le récepteur influant essentiellement sur les opérations de filtrage, les conditions de réception peuvent être établies à l'aide d'un formalisme similaire. Dans cette section, nous distinguerons donc principalement la détection pour les modulations M-aire et les modulations CDMA.

8.2.2.1 Cas des modulations M-aire

Considérons la transmission du symbole k pendant le l-ième temps symbole. Dans le cas du filtre blanchissant, nous avons établi section 5.2.1.3 que le symbole détecté $\hat{k}$ est donné par :

$$\hat{k} = \arg \max_{m \in [\![0, M-1]\!]} \hat{\mathbf{v}}^t \hat{\mathbf{d}}_m, \tag{8.2}$$

où $\hat{\mathbf{v}}$ est l'observation du signal reçu après blanchiment sur le l-ième temps symbole en amont du corrélateur et $\hat{\mathbf{d}}_m$ est le m-ième vecteur du dictionnaire de réception $\hat{\mathcal{D}}$. k est correctement détecté si la corrélation entre le signal reçu et le vecteur du dictionnaire de réception associé à k est supérieure aux corrélations avec les autres vecteurs du dictionnaire de réception. Ce formalisme est identique dans le cas de l'égaliseur à condition que le dictionnaire de réception soit le dictionnaire d'émission $\mathcal{D}$.

3. Cette insertion est effectuée sur une durée N_s de l'ordre d'une centaine d'échantillons alors que l'actualisation de α est effectuée toutes les $N_m = 1$ secondes. α est donc constant pendant l'insertion de k.

La copie locale du récepteur permet d'estimer le signal reçu sous la forme de $\tilde{\mathbf{x}} + \tilde{\mathbf{t}} + \tilde{\mathbf{p}}$, somme de la composante audio blanchie ou "égalisée" $\tilde{\mathbf{x}}$, du tatouage filtré $\tilde{\mathbf{t}}$ et d'un bruit de canal additif $\tilde{\mathbf{p}}$. Elle fournit également une estimation du dictionnaire de réception : $\tilde{\mathcal{D}}$ dans le cas du filtrage blanchissant et $\mathcal{D}$ pour l'égaliseur. Les conditions d'une détection correcte et robuste de k peuvent donc être établies sous la forme de $M-1$ inéqualités :

$$\forall m \neq k, \left(\tilde{\mathbf{x}} + \tilde{\mathbf{t}} + \tilde{\mathbf{p}}\right)^t \breve{\mathbf{d}}_k \geq \left(\tilde{\mathbf{x}} + \tilde{\mathbf{t}} + \tilde{\mathbf{p}}\right)^t \breve{\mathbf{d}}_m, \tag{8.3}$$

où $\{\breve{\mathbf{d}}_m\}_{m=0..M-1}$ désigne les vecteurs de l'estimation du dictionnaire de réception $\breve{\mathcal{D}}$ (c'est à dire $\tilde{\mathcal{D}}$ pour le récepteur par filtre blanchissant et $\mathcal{D}$ pour l'égaliseur).

8.2.2.2 Cas des modulations CDMA

Considérons maintenant la transmission du K-uplet binaire $\mathbf{b}$ par une modulation CDMA. La détection est un démodulateur sous-optimal par décorrélation ; il consiste à projeter le signal reçu $\hat{\mathbf{v}}$ après filtrage sur l'ensemble des vecteurs du dictionnaire de réception ($\hat{\mathcal{D}}$ dans le cas du récepteur par filtrage blanchissant et $\mathcal{D}$ dans le cas de l'égaliseur de Wiener) puis à pondérer le vecteur des projections par la matrice inverse d'autocorrélation empirique du dictionnaire de réception : cette matrice est $\mathcal{R}_{\hat{\mathcal{D}}} = \hat{\mathcal{D}}^t\hat{\mathcal{D}}$ dans le cas du récepteur par filtre blanchissant et $\mathcal{R}_{\mathcal{D}} = \mathcal{D}^t\mathcal{D}$ pour l'égaliseur. La décision tient alors au signe de chaque composante du vecteur projeté $\hat{\mathbf{c}}'$ ainsi obtenu. Le K-uplet $\hat{\mathbf{b}}$ détecté est donc donné par la relation :

$$\hat{\mathbf{b}} = \frac{\hat{\mathbf{a}} + 1}{2} \text{ avec } \hat{\mathbf{a}} = \text{sign}\left(\hat{\mathbf{c}}'\right) \tag{8.4}$$

Comme dans le cas de modulation M-aire, la copie locale du récepteur à l'émetteur permet d'estimer le signal reçu à l'entrée du corrélateur sous la forme $\tilde{\mathbf{x}} + \tilde{\mathbf{t}} + \tilde{\mathbf{p}}$, somme de la composante audio blanchie ou égalisée $\tilde{\mathbf{x}}$, du tatouage filtré $\tilde{\mathbf{t}}$ et d'un bruit de canal additif $\tilde{\mathbf{p}}$. Elle fournit là encore une estimation du dictionnaire de réception $\breve{\mathcal{D}}$ (correspondant à $\tilde{\mathcal{D}}$ pour le filtre blanchissant et $\mathcal{D}$ pour l'égaliseur). La détection peut donc être estimée en fonction du signe des composantes de $\tilde{\mathbf{c}}' = \mathcal{R}_{\breve{\mathcal{D}}}^{-1}\breve{\mathcal{D}}^t\left(\tilde{\mathbf{x}} + \tilde{\mathbf{t}} + \tilde{\mathbf{p}}\right)$.

La détection du K-uplet binaire $\mathbf{b}$ émis est correcte si le signe des composantes de $\tilde{\mathbf{c}}'$ est identique au signe des composantes de $2\mathbf{b}-1$. Autrement dit, en notant $\tilde{\mathbf{c}}'[m]$ la m-ième composante de $\tilde{\mathbf{c}}'$ et $b[m]$ la m-ième composante de $\mathbf{b}$, la détection de $b[m]$ est correcte si $(2b[m]-1)\tilde{\mathbf{c}}'[m] > 0$. Les conditions de la détection correcte de $\mathbf{b}$ peut donc être exprimée sous la forme vectorielle suivante :

$$(2\mathbf{b} - 1) \otimes \left(\mathcal{R}_{\breve{\mathcal{D}}}^{-1}\breve{\mathcal{D}}^t\left(\tilde{\mathbf{x}} + \tilde{\mathbf{t}} + \tilde{\mathbf{p}}\right)\right) > 0, \tag{8.5}$$

où $\otimes$ désigne le produit terme à terme et la fonction $x > 0$ est généralisée au cas vectoriel (signifiant que toutes les composantes de $(2\mathbf{b} - 1) \otimes \tilde{\mathbf{c}}$ doivent être positives).

Conclusion

L'objectif de la stratégie d'insertion informée consiste donc à rechercher le tatouage $\mathbf{t}$ tel que :

- le signal modulé $\mathbf{v}$ dont il est issu satisfasse à la contrainte d'inaudibilité (8.1),
- sa version filtrée $\tilde{\mathbf{t}}$ (par les filtres de la chaîne) soit solution des conditions de détection (8.3), et (8.5), choisies en fonction de la configuration de modulation et de réception, pour un bruit de canal $\tilde{\mathbf{p}}$ de puissance maximale.

8.2.3 Expression du signal modulé recherché

La conception même du système (utilisant des filtrages et projetant le signal reçu sur le dictionnaire de réception) permet de rechercher le signal modulé associé au tatouage solution sous la forme d'une combinaison linéaire des vecteurs du dictionnaire :

$$\mathbf{v} = \sum_{m=0}^{M} \lambda_m \mathbf{d}_m = \mathcal{D}\lambda, \tag{8.6}$$

où $\lambda = [\lambda_0 ... \lambda_{M-1}]^t$ sont les coefficients de la combinaison linéaire, représentés sous forme vectorielle. Cette section s'intéresse à la démonstration de cette assertion.

8.2.3.1 Cas des modulations M-aire

Les inégalités (8.3) traduisant la transmission correcte et robuste de l'information k montrent que seule la projection du signal reçu dans l'espace engendré par les vecteurs du dictionnaire de réception ($\tilde{\mathcal{D}}$ ou $\mathcal{D}$ en fonction du récepteur) intervient dans le procédé de détection : cet espace est l'***espace de détection***. De fait, il suffit de rechercher le tatouage filtré $\tilde{\mathbf{t}}$ et de manière équivalente le signal modulé filtré $\tilde{\mathbf{v}} = \frac{1}{\alpha}\tilde{\mathbf{t}}$ dans cet espace ; en effet, toutes composantes de ces signaux en dehors de l'espace de détection sont supprimées par le corrélateur et n'interviennent donc pas dans le processus de décision.

Le signal modulé filtré $\tilde{\mathbf{v}}$ peut finalement être exprimé sous la forme d'une combinaison linéaire des vecteurs du dictionnaire de réception, dont les coefficients sont notés sous forme vectorielle $\lambda = [\lambda_0 ... \lambda_{M-1}]^t$. Cette combinaison prend la forme :

- dans le cas du récepteur par filtre blanchissant :

$$\tilde{\mathbf{v}} = \sum_{m=0}^{M-1} \lambda_m \tilde{\mathbf{d}}_m = \tilde{\mathcal{D}}\lambda. \tag{8.7}$$

  D'après le schéma de la copie locale du récepteur à l'émetteur, $\tilde{\mathbf{v}}$ tout comme $\tilde{\mathcal{D}}$ résultent du filtrage de $\mathbf{v}$ et de $\mathcal{D}$ par $H(f)$ et $G(f)$. Le filtrage étant linéaire, nous en déduisons de 8.7 que $\mathbf{v}$ peut s'exprimer comme une combinaison linéaire des vecteurs du dictionnaire d'émission $\mathcal{D}$ de même coefficients λ.

- dans le cas de l'égaliseur :

$$\tilde{\mathbf{v}} = \sum_{m=0}^{M-1} \gamma_m \mathbf{d}_m = \mathcal{D}\gamma, \tag{8.8}$$

  puisque la réception exploite directement le dictionnaire d'émission.
  $\tilde{\mathbf{v}}$ étant par construction du filtre de Wiener une estimation du signal modulé $\mathbf{v}$, il paraît alors judicieux de rechercher $\mathbf{v}$ sous la forme d'une combinaison linéaire des vecteurs du dictionnaire d'émission : $\mathbf{v} = \mathcal{D}\lambda$. La linéarité du filtrage établit alors $\tilde{\mathbf{v}} = \tilde{\mathcal{D}}\lambda$, où $\tilde{\mathcal{D}}$ est par hypothèse l'ensemble des vecteurs du dictionnaire d'émission filtrés par le filtre de Wiener. Les coefficients de γ sont alors déduits en projetant $\tilde{\mathcal{D}}\lambda$ sur l'espace de détection engendré par le dictionnaire de réception $\mathcal{D}$, à savoir : $\gamma = \mathcal{D}^t\tilde{\mathcal{D}}\lambda$.

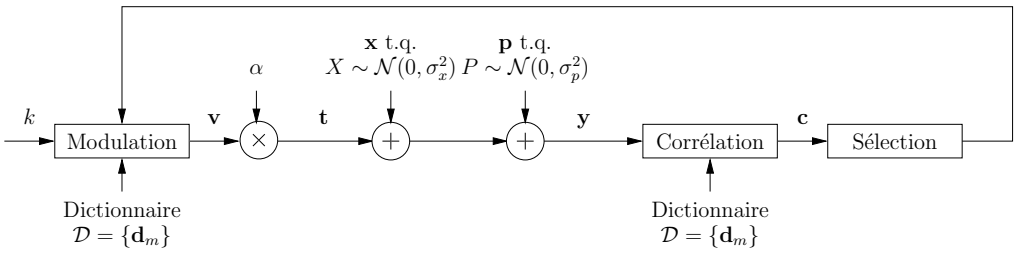


Figure 8.3 – Modèle de la stratégie d'insertion informée pour la maximisation de la robustesse sous contrainte d'inaudibilité.

8.2.3.2 Cas des modulations CDMA

Dans le cas des modulations CDMA, rechercher le signal modulé sous la forme d'une combinaison linéaire des vecteurs du dictionnaire est inhérent à la définition des modulations CDMA. En effet, le K-uplet $\mathbf{b} = [b[0], ..., b[K-1]]^t$ est transmis via le signal modulé :

$$\mathbf{v} = \sum_{m=0}^{M-1} \gamma_m (2b[m] - 1)\mathbf{d}_m, \tag{8.9}$$

où γ_m est par définition l'énergie dédiée à la transmission du m-ième bit du K-uplet.

Conclusion

Exprimer le signal modulé $\mathbf{v}$ sous la forme d'une combinaison linéaire des vecteurs du dictionnaire d'émission s'avère particulièrement intéressant. En effet, il est alors possible de travailler indifféremment sur $\mathbf{v}$, $\tilde{\mathbf{v}}$ ou $\tilde{\mathbf{t}}$. En effet, d'après les schéma des copies locales de récepteur à l'émetteur, nous connaissons la relation de filtrage liant $\mathbf{v}$ et $\tilde{\mathbf{v}}$, celle liant $\mathcal{D}$ et $\tilde{\mathcal{D}}$ ainsi que l'expression $\tilde{\mathbf{t}} = \alpha\tilde{\mathbf{v}}$ liant le tatouage filtré au signal modulé filtré. Connaissant les coefficients de la combinaison linéaire $\mathcal{D}\lambda$, les trois signaux se déduisent donc les uns des autres par la relation suivante :

$$\mathbf{v} = \mathcal{D}\lambda \Longleftrightarrow \tilde{\mathbf{v}} = \tilde{\mathcal{D}}\lambda \Longleftrightarrow \tilde{\mathbf{t}} = \alpha\tilde{\mathcal{D}}\lambda. \tag{8.10}$$

La stratégie d'insertion informée peut désormais s'exprimer non plus comme la recherche du signal de tatouage $\mathbf{t}$ adapté mais comme celle des coefficients λ des combinaisons linéaires tel que :

- le signal modulé $\mathbf{v} = \mathcal{D}\lambda$ satisfait à la contrainte d'inaudibilité (8.1),
- le signal de tatouage filtré $\tilde{\mathbf{t}} = \alpha\lambda\tilde{\mathcal{D}}$ soit solution des conditions de détection (8.3) ou (8.5), choisies en fonction de la configuration de modulation et de réception, pour un bruit de canal $\tilde{\mathbf{p}}$ de puissance maximale.

8.3 Modèle de la stratégie d'insertion informée

Les similitudes dans le formalisme des stratégies d'insertion informée proposées pour les deux récepteurs conduisent à proposer un modèle de la stratégie d'insertion informée unifié : le signal modulé $\tilde{\mathbf{v}}$ en sortie des filtres de la chaîne est recherché sous la forme d'une combinaison linéaire

des vecteurs d'un dictionnaire $\tilde{\mathcal{D}}$. Il peut donc être vu comme le résultat d'une modulation exploitant le dictionnaire $\tilde{\mathcal{D}}$; le signal reçu en entrée du corrélateur est la somme d'une composante provenant du signal audio $\tilde{\mathbf{x}}$, du tatouage $\tilde{\mathbf{t}} = \alpha\tilde{\mathbf{v}}$ et d'un bruit de canal $\tilde{\mathbf{p}}$. La décision tient au calcul de la corrélation $\tilde{\mathbf{c}}$ entre le signal reçu et les vecteurs du dictionnaire de réception.

Le modèle de la stratégie informée, présenté figure 8.3, est donc identique à celui établi pour l'analyse des performances théoriques de la chaîne de tatouage non informée, proposé figure 4.1, à deux exceptions près : cette fois le signal audio $\mathbf{x}$ est supposé connu lors de l'étape de modulation et un bruit de canal $\mathbf{p}$ est ajouté.

Ce modèle nous permettra d'établir les différentes stratégies d'insertion informée indépendamment de la configuration de réception (filtre blanchissant ou égaliseur de Larbi) avant de les appliquer au système de tatouage. Ce modèle reste dépendant de la modulation choisie. De fait, dans le second chapitre de cette troisième partie, nous nous intéresserons aux modulations M-aire, tandis que les modulations CDMA feront l'objet du troisième chapitre.

Chapitre 9

Tatouage M-aire de robustesse maximale aux perturbations

L'objectif de ce chapitre est d'établir une stratégie d'insertion informée permettant de choisir un signal de tatouage $\mathbf{t}$ qui satisfait l'inaudibilité de la transmission et maximisant sa robustesse aux perturbations lorsque l'émetteur exploite une modulation M-aire.

9.1 De la nécessité d'une stratégie informée

9.1.1 Interprétation géométrique de la stratégie d'insertion

Reprenons dès à présent l'interprétation géométrique de la stratégie d'insertion proposée par Miller [MCB00] dans le cadre du modèle de la chaîne établi section 8.3. Le tatouage, porteur de l'information k, doit être choisi de sorte à concilier deux contraintes : la contrainte d'inaudibilité et la contrainte de détection correcte et robuste de k. Ces deux contraintes peuvent être représentées géométriquement dans l'espace de détection. Cet espace est l'espace engendré par les vecteurs du dictionnaire. Dans cet espace, les contraintes définissent deux régions d'intérêt, dépendantes du signal audio $\mathbf{x}$:

- la ***région d'inaudibilité*** $\mathfrak{R}_i$ (ou de distorsion acceptable). Cette région est l'ensemble des signaux audio tatoués $\mathbf{y}$ perceptuellement proche du signal audio original $\mathbf{x}$.
  L'inaudibilité du tatouage est garantie par la contrainte de puissance imposée au signal modulé $\mathbf{v}$ dont est issu le tatouage $\mathbf{t} = \alpha\mathbf{v}$: ce signal doit être de puissance unité. La région de distorsion acceptable est donc définie par :

$$\mathfrak{R}_i = \left\{ \mathbf{y} = \mathbf{x} + \mathbf{t} \,/\, \sigma_t^2 = \frac{\mathbf{t}^t\mathbf{t}}{N_s} \leq \alpha^2 \right\}. \tag{9.1}$$

  Cette région prend la forme d'une hyper-sphère centrée sur le signal audio $\mathbf{x}$.

- la ***région de détection*** $\mathfrak{R}_d$, ensemble des signaux audio tatoués pour lesquels l'information k est correctement détectée. Cette région peut être restreinte à un sous-ensemble appelé ***région de détection robuste*** pour laquelle la détection de l'information k est robuste à l'ajout d'un bruit de canal.

La décision est basée sur le résultat de la corrélation $\mathbf{c}$ entre le signal reçu (le bruit audio $\mathbf{x}$ auquel s'ajoute le tatouage $\mathbf{t}$ et un éventuel bruit de canal) et le dictionnaire. Le corrélateur décide que l'information k a été reçue si la corrélation du signal reçu avec le k-ième vecteur du dictionnaire $\mathbf{d}_k$ est supérieure aux corrélations du signal reçu avec les autres vecteurs du dictionnaire. En l'absence de bruit de canal, la région de détection de l'information k est donc établie par :

$$\mathfrak{R}_d = \left\{ \mathbf{y} = \mathbf{x} + \mathbf{t} \,/\, \forall m \neq k, (\mathbf{x} + \mathbf{t})^t \mathbf{d}_k > (\mathbf{x} + \mathbf{t})^t \mathbf{d}_m \right\}. \tag{9.2}$$

Il s'agit d'un hyper-cône centré sur $\mathbf{d}_k$.

Le tatouage $\mathbf{t}$ doit finalement être choisi de sorte que le signal audio tatoué qui résulte de son insertion $\mathbf{y} = \mathbf{x} + \mathbf{t}$ soit à l'intersection de ces deux régions (à condition que cette intersection ne soit pas vide). La façon de choisir ce tatouage définit la stratégie d'insertion.

9.1.2 Un exemple en dimension 2

Les figures 9.1 visualisent les deux régions d'intérêt dans le cas d'une modulation orthogonale utilisant 2 vecteurs et pour différentes configurations du signal $\mathbf{x}$. Nous supposerons que l'information à tatouer est le bit $'1'$. Le tatouage choisi pour moduler l'information $'1'$ est le vecteur du dictionnaire $\mathbf{d}_1$ pondéré par le facteur d'amplitude α. Pour un espace signal à deux dimensions, la région de distorsion acceptable est un cercle de rayon $\alpha||\mathbf{d}_1||$ centré sur le signal audio $\mathbf{x}$. La région de détection est l'ensemble de points dont la corrélation avec $\mathbf{d}_1$ est plus grande que celle avec $\mathbf{d}_0$. Or,

$$\mathbf{y}^t\mathbf{d}_1 > \mathbf{y}^t\mathbf{d}_0 \Leftrightarrow \mathbf{y}^t(\mathbf{d}_1 - \mathbf{d}_0) > 0. \tag{9.3}$$

Cette région est donc le demi-plan limité par la droite dirigée par $\mathbf{d}_1 + \mathbf{d}_0$ et contenant $\mathbf{d}_1$ (sur les figures, le demi-plan supérieur). Cette droite marque la frontière entre la région de détection de l'information $'1'$ et celle de $'0'$. Un signal de tatouage $\mathbf{t}$, pour être adapté, doit donc être choisi de sorte que $\mathbf{y} = \mathbf{x} + \mathbf{t}$ soit sur l'arc de cercle définissant l'intersection entre la région de distorsion acceptable et la région de détection, à condition bien sûr que cet arc de cercle ne soit pas vide.

Configuration très favorable

La première configuration envisagée, donnée figure 9.1 (a), est celle où le signal audio $\mathbf{x}$ se trouve dans la région de détection. Le signal audio est plus corrélé avec $\mathbf{d}_1$ qu'avec $\mathbf{d}_0$: il est donc à lui seul porteur de l'information à tatouer $'1'$. L'intersection entre la région de détection et la région de distorsion acceptable n'est pas vide. Le signal de tatouage $t = \alpha\mathbf{d}_1$ issu de la modulation du bit $'1'$ satisfait donc aux contraintes de transmission. On pourra également remarquer que dans ce contexte précis, l'ajout d'un signal de tatouage n'est pas nécessaire. En effet, en choisissant pour signal audio tatoué $\mathbf{y}$ le signal audio original $\mathbf{x}$ (avec ici $\mathbf{t} = 0$), aucune dégradation n'est introduite et la détection correcte de l'information peut être obtenue puisque $\mathbf{x}$ est porteur à lui seul de l'information.

Configurations favorables

La seconde et la troisième configuration, donnée figures 9.1 (b) et (c), présentent le cas d'un signal audio hors de la région de détection mais pour lequel l'intersection entre les 2 régions d'intérêt n'est pas vide. $\mathbf{x}$ étant plus fortement corrélé avec $\mathbf{d}_0$ qu'avec $\mathbf{d}_1$, transmettre $\mathbf{x}$ seul

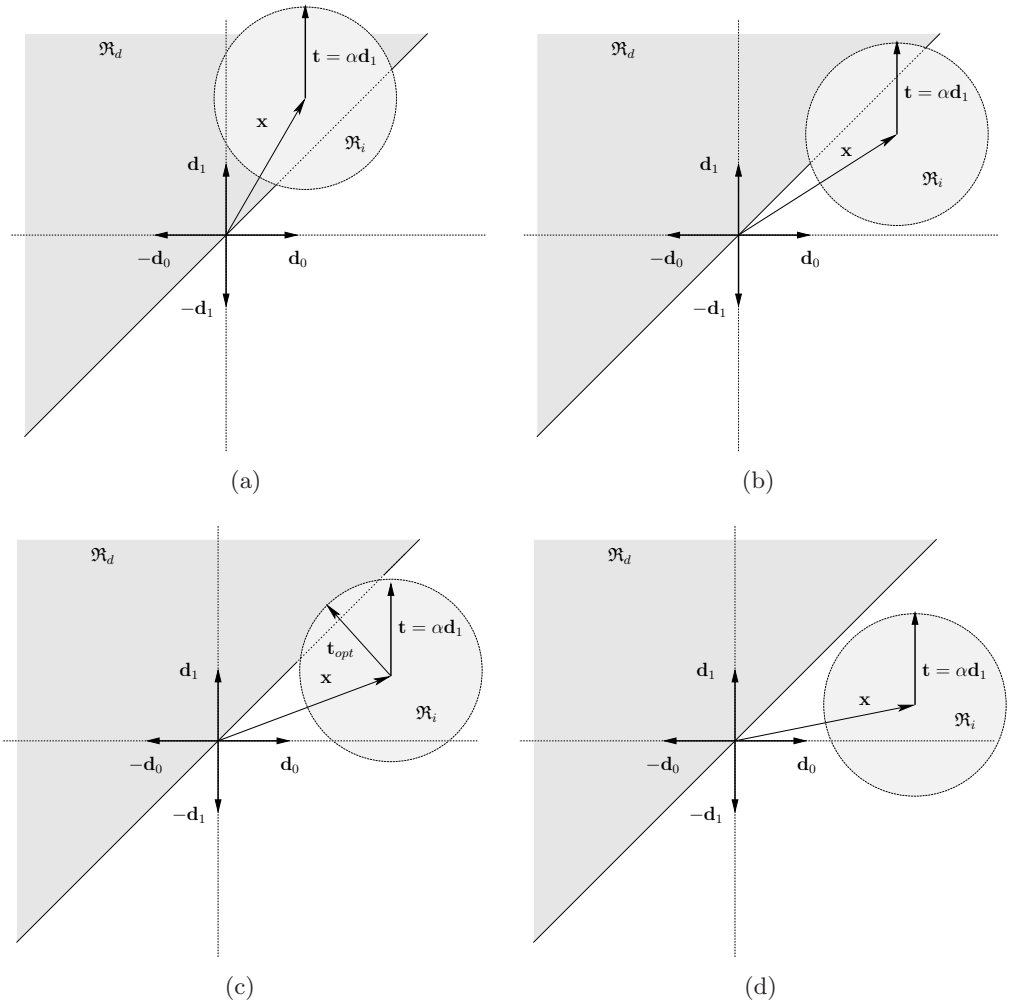


Figure 9.1 – Mise en évidence des régions de distorsion acceptable et de détection pour différentes configurations de signaux de la plus favorable à la moins favorable

conduirait à une erreur de détection. L'ajout d'un signal de tatouage devient nécessaire. Dans le cas de la configuration (b), le signal de tatouage habituel $\mathbf{t} = \alpha\mathbf{d}_1$ permet d'obtenir un signal audio tatoué $\mathbf{y} = \mathbf{x} + \mathbf{t}$ à l'intérieur de la région de détection. Les conditions de transmission sont donc satisfaites. Par contre, dans le cas de la configuration (c), le tatouage $\mathbf{t} = \alpha\mathbf{d}_1$ ne permet plus d'obtenir une transmission sans erreur puisque le signal audio tatoué résultant est en dehors de la région de détection. Pourtant, l'intersection entre les 2 régions n'est pas vide. Il existe donc plusieurs alternatives au signal de tatouage $\mathbf{t} = \alpha\mathbf{d}_1$ qui satisfont les conditions d'une transmission correcte : l'une d'elle est le signal t_{opt}, présenté sur la figure, dont les caractéristiques seront détaillées par la suite.

Configuration défavorable

La dernière configuration, donnée figure 9.1 (d), présente le contexte de transmission le plus défavorable. Le signal audio $\mathbf{x}$ est ici très fortement corrélé avec le vecteur $\mathbf{d}_0$ à tel point que l'intersection entre les deux régions d'intérêt est vide. Dans ce cas, aucun signal de tatouage (de

distorsion acceptable) ne permet d'obtenir les conditions d'une transmission correcte. Les deux régions sont ici trop éloignées l'une de l'autre (au regard de la contrainte de puissance imposée au tatouage). Une alternative doit donc être envisagée pour minimiser l'écart maximum existant entre les 2 régions. Cette alternative peut s'appuyer sur la stratégie d'insertion proposée par Costa [Cos83]. Cette stratégie vise notamment à démultiplier les régions de détection associées à l'information à transmettre pour garantir la proximité de l'espace de distorsion acceptable et de l'espace de détection par l'utilisation d'un dictionnaire structuré.

9.1.3 Conséquences sur la stratégie d'insertion

Cette interprétation géométrique révèle la nécessité de modifier la stratégie d'insertion pour tendre vers une stratégie informée, c'est à dire exploitant la connaissance *a priori* du signal audio à l'émetteur. Cette stratégie doit permettre de choisir une région de détection proche du signal audio garantissant l'existence d'au moins un tatouage satisfaisant les conditions de transmission. Dans cet objectif, le dictionnaire peut être modifié pour adopter la structure proposée par Costa. L'étude des performances aussi bien théoriques que pratiques fait l'objet de la section 9.2. A supposer qu'une région de détection proche du signal audio existe, plusieurs tatouages peuvent être solutions. Parmi ces solutions, il convient alors de choisir celle qui minimisera la probabilité d'erreur de transmission pour un bruit de canal maximum. La section 9.3 propose différentes stratégies d'insertion maximisant la robustesse de la transmission sous contrainte d'inaudibilité ; ces stratégies donnent lieu à différents problèmes d'optimisation que nous nous proposons de résoudre par des algorithmes d'optimisation adaptés. Les performances de ces stratégies seront comparées sur le modèle du système avant d'être mise en œuvre sur le système pratique en boucle fermée. Pour finir, le dictionnaire structuré et les procédures d'optimisation sont combinées donnant lieu au schéma final du système de tatouage informé en boucle fermée. Les performances pratiques de ce système seront finalement évaluées.

9.2 Vers un dictionnaire structuré

Dans un premier temps, nous modifions le dictionnaire du système de tatouage de sorte à lui conférer la structure proposée par le schéma de Costa [Cos83] : au lieu de ne proposer qu'un seul vecteur pour transmettre un symbole d'information, le dictionnaire offre plusieurs vecteurs ; la fonction d'insertion doit alors sélectionner le vecteur le plus susceptible de transmettre l'information sans erreur.

Les résultats théoriques de Costa [1] sur la capacité du canal audio aussi bien que les TEBs obtenus par Miller [Mil01] sur la mise en œuvre pratique du dictionnaire structuré pour des images ont déjà établi l'efficacité de ce choix de dictionnaire : à supposer le dictionnaire de canal suffisamment grand, la probabilité d'erreur de transmission peut être rendue aussi petite que l'on souhaite à condition que le débit de transmission ne dépasse pas la capacité d'insertion de Costa $C_i = \frac{1}{2}\log_2\left(1+\frac{\sigma_t^2}{\sigma_p^2}\right)$. Notre système de tatouage, de part sa construction, ne peut mettre en œuvre qu'un dictionnaire de canal de dimension relativement faible ; en effet, la réception par filtrage blanchissant nécessite notamment l'estimation d'un dictionnaire de réception pour chaque symbole émis, résultant du filtrage du dictionnaire d'émission. Le coût du récepteur devient donc prohibitif si le dictionnaire utilisé atteint les dimensions escomptées du schéma de

1. Ces résultats ont été rappelés section 3.3.1.

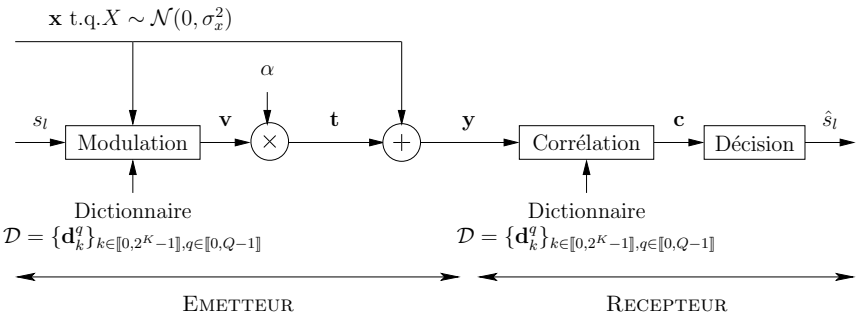


Figure 9.2 – Schéma du système de tatouage.

Costa. Nous avons donc souhaité évaluer l'**efficacité d'un dictionnaire structuré de faible dimension** sur les performances du système de tatouage. Dans cet objectif, nous proposons l'évaluation des performances théoriques du système sous la forme d'un calcul de probabilité d'erreur, qui fait défaut dans la littérature. Cette évaluation fait l'objet de la prochaine section. La mise en œuvre du dictionnaire structuré est ensuite appliquée au système de tatouage en boucle fermée pour jauger de ses performances sur des signaux réels.

9.2.1 Etude du modèle informé du système de tatouage

Reprenons le modèle du système de tatouage développé précédemment pour le calcul des performances théoriques de la chaîne de tatouage dans le cas d'une modulation M-aire, en exploitant cette fois la connaissance du signal audio $\mathbf{x}$ à l'émetteur. Le modèle développé est présenté figure 9.2.

Comme proposé par le schéma de Costa, l'étape de modulation est modifiée de sorte à multiplier les régions de détection associées à une même information. Précédemment, le dictionnaire d'émission $\mathcal{D}$ contenait autant de signaux que de valeurs possibles du symbole à émettre s de valeur k : la modulation consistait donc à associer de manière bijective l'information k au k-ième vecteur du dictionnaire.

Le dictionnaire $\mathcal{D}$ est maintenant structuré sous la forme d'un ensemble de 2^K sous-dictionnaires, noté $\mathcal{D}_k$ pour $k \in [\![0, 2^K - 1]\!]$. Chaque sous-dictionnaire $\mathcal{D}_k$ contient Q signaux notés $\mathbf{d}_k^q$ pour $q \in [\![0, Q-1]\!]$ choisis blancs, de puissance unité, et étalés spectralement dans la bande de fréquence $[0, F_c]$:

$$\mathcal{D}_k = \left\{\mathbf{d}_k^q\right\}_{q\in[\![0,Q-1]\!]}. \tag{9.4}$$

Le dictionnaire $\mathcal{D}$ est donc étendu à $M = Q2^K$ vecteurs.

Chaque signal $\mathbf{d}_k^q$ du sous-dictionnaire $\mathcal{D}_k$ représente l'information k. L'étape de modulation doit donc permettre de choisir parmi les Q signaux de $\mathcal{D}_k$ porteur de l'information k celui qui sera le plus adapté à la transmission de k. Ce dernier sera noté $\mathbf{d}_k^{opt}$. La modulation consiste désormais à associer de manière surjective l'information k au vecteur $\mathbf{d}_k^{opt}$. Le signal de tatouage s'écrit alors :

$$\mathbf{t} = \alpha \mathbf{d}_k^{opt}. \tag{9.5}$$

9.2.1.1 Conditions de détection

Étant données les modifications introduites sur le dictionnaire, k est correctement détectée s'il existe au moins un signal du sous-dictionnaire $\mathcal{D}_k$ dont la corrélation avec le signal reçu est supérieure aux corrélations entre le signal reçu et les vecteurs des autres sous-dictionnaires. Cette condition se traduit par l'inégalité suivante :

$$\exists q \in [\![0, Q-1]\!], \forall m \neq k, \forall i \in [\![0, Q-1]\!], (\mathbf{x}+\mathbf{t}+\mathbf{p})^t \mathbf{d}_k^q > (\mathbf{x}+\mathbf{t}+\mathbf{p})^t \mathbf{d}_m^i. \tag{9.6}$$

Le signal du sous-dictionnaire $\mathcal{D}_k$ le plus susceptible de satisfaire cette inégalité est naturellement celui dont la corrélation avec le signal reçu est la plus grande :

$$q = \arg \max_{p \in [\![0,Q-1]\!]} (\mathbf{x}+\mathbf{t}+\mathbf{p})^t \mathbf{d}_k^p. \tag{9.7}$$

9.2.1.2 Choix des sous-dictionnaires

Plusieurs configurations du dictionnaire peuvent être envisagées. La théorie des communications numériques [Pro01] préconise d'utiliser des signaux orthogonaux ou antipodaux (car maximisant les distances entre les vecteurs dans l'espace de détection). Trois configurations, illustrées figure 9.3, peuvent alors être déduites :

Dictionnaire orthogonal

Pour cette configuration, tous les signaux contenus dans le dictionnaire sont orthogonaux les uns aux autres. Chaque sous-dictionnaire engendre un espace de dimension Q, orthogonal à ceux engendrés par les autres sous-dictionnaires. L'ensemble du dictionnaire définit finalement un espace de dimension $Q2^K$. Ainsi :

$$\begin{aligned} \mathcal{D} = &\left\{ \mathcal{D}_m = \{\mathbf{d}_m^q\}_{q \in [\![0,Q-1]\!]} \right\}_{m \in [\![0,2^K-1]\!]} \\ &\text{tel que } \forall k,l,p,q, \frac{1}{N_s}\, \mathbf{d}_k^{p\,t}\, \mathbf{d}_l^q = \delta(k,l)\delta(p,q). \end{aligned} \tag{9.8}$$

Sous-dictionnaires biorthogonaux

Chaque sous-dictionnaire est ici constitué de Q signaux biorthogonaux. Ces signaux définissent un espace de dimension $\dfrac{Q}{2}$ et sont choisis de sorte à être orthogonal aux espaces engendrés par les autres sous-dictionnaires. L'ensemble du dictionnaire définit finalement un espace de dimension $Q2^{K-1}$. Ainsi :

$$\begin{aligned} \mathcal{D} = &\left\{ \mathcal{D}_m = \{\mathbf{d}_m^q\}_{q \in [\![0,Q-1]\!]} \right\}_{m \in [\![0,2^K-1]\!]} \\ &\text{tel que } \begin{cases} \forall k, \forall p,q \in [\![0, \frac{Q}{2}-1]\!], \frac{1}{N_s}\, \mathbf{d}_k^{p\,t}\, \mathbf{d}_k^q = \delta(p,q) \\ \forall k, \forall q \in [\![0, \frac{Q}{2}-1]\!], \mathbf{d}_k^{q+\frac{Q}{2}} = -\mathbf{d}_k^q \\ \forall k \neq l, \forall p,q, \frac{1}{N_s}\, \mathbf{d}_k^{p\,t}\, \mathbf{d}_l^q = 0 \end{cases} . \end{aligned} \tag{9.9}$$

Sous-dictionnaires orthogonaux

Cette configuration peut être vue comme l'extension de la modulation biorthogonale. Les sous-dictionnaires sont constitués de signaux orthogonaux définissant chacun un espace de Q dimensions. Le dictionnaire peut être défini par la seule donnée de la moitié des sous-dictionnaires : les 2^{K-1} premiers sous-dictionnaires contiennent des signaux orthogonaux

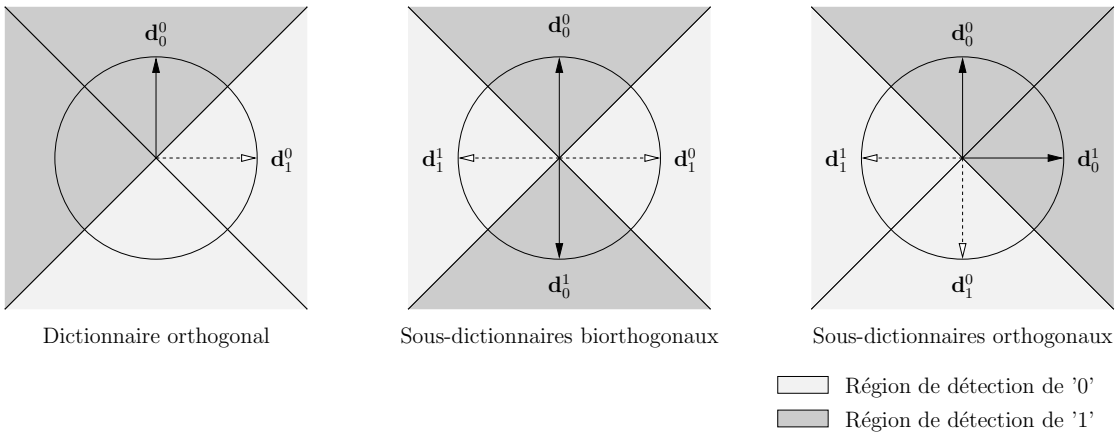


Figure 9.3 – Configurations des constellations engendrées par le dictionnaire pour une même taille de l'espace signal lorsque $K = 1$.

définissant des espaces orthogonaux entre eux. La seconde moitié des sous-dictionnaires est égale à l'opposée de la première. Comme dans le cas précédent, l'ensemble du dictionnaire définit un espace de dimension $Q2^{K-1}$, avec :

$$\mathcal{D} = \left\{\mathcal{D}_m = \{\mathbf{d}_m^q\}_{q\in[\![0,Q-1]\!]}\right\}_{m\in[\![0,2^K-1]\!]}$$
$$\text{tel que } \begin{cases} \forall k \in [\![0, 2^{K-1}-1]\!], \forall q, \mathbf{d}_{k+2^{K-1}}^q = -\mathbf{d}_k^q \\ \forall k, \forall p \neq q, \frac{1}{N_s}\, \mathbf{d}_k^{p\,t}\, \mathbf{d}_k^q = 0 \end{cases} . \tag{9.10}$$

La figure 9.3 présente les trois configurations lorsque l'espace engendré par les vecteurs du dictionnaire est limité à la dimension 2 et dans le cas d'une transmission binaire ($K = 1$). Elle met également en évidence les régions de détection associées à l'information $'0'$ et à l'information $'1'$. Dans ce contexte, la configuration du dictionnaire orthogonal se réduit au cas où $Q = 1$: chaque sous-dictionnaire contient un unique vecteur. Les régions de détection sont deux demi-plans séparés par la somme des deux vecteurs. Pour les deux autres configurations, $Q = 2$: chaque sous-dictionnaire contient 2 vecteurs représentant une même valeur de l'information émise. Les régions de détection sont la fusion des cônes de détection centrés sur chacun des vecteurs représentant la même information. Pour la configuration sous-dictionnaires orthogonaux, ces régions s'unissent pour donner un demi-plan dirigé par la somme des deux vecteurs ; il engendre donc un pavage de l'espace similaire à celui d'un dictionnaire orthogonal. Dans le cas de la configuration sous-dictionnaire biorthogonaux, le choix des vecteurs permet d'obtenir un pavage de l'espace signal en 4 régions alternées.

Par la suite, on ne s'intéressera qu'aux deux premiers cas de figure : celui du dictionnaire orthogonal et des sous-dictionnaires biorthogonaux.

9.2.1.3 Choix du signal du dictionnaire le plus adapté

Le signal du dictionnaire $\mathbf{d}_k^{opt}$ le plus adapté à la transmission de l'information k est celui qui maximise les chances de détecter correctement l'information émise. Il doit donc respecter les conditions de détection précisées par l'inégalité (9.6) pour $\mathbf{t} = \alpha\mathbf{d}_k^{opt}$:

$$\forall m \neq k, \forall i, (\mathbf{x}+\alpha\mathbf{d}_k^{opt}+\mathbf{p})^t\mathbf{d}_k^q > (\mathbf{x}+\alpha\mathbf{d}_k^{opt}+\mathbf{p})^t\mathbf{d}_m^i \text{ avec } q = \arg\max_{p\in[\![0,Q-1]\!]}(\mathbf{x}+\mathbf{t}+\mathbf{p})^t\mathbf{d}_k^p \tag{9.11}$$

La détection est donc d'autant plus aisée que la plus grande corrélation entre le signal reçu et les vecteurs du sous-dictionnaire $\mathcal{D}_k$, $\max_{p\in[\![0,Q-1]\!]}(\mathbf{x}+\mathbf{t}+\mathbf{p})^t\mathbf{d}_k^p$, est grande. Le signal $\mathbf{d}_k^{opt}$ est donc choisi de sorte que :

$$opt = \arg\max_{p\in[\![0,Q-1]\!]}(\mathbf{x}+\alpha\mathbf{d}_k^p+\mathbf{p})^t\mathbf{d}_k^q, \text{ avec } q = \arg\max_{p\in[\![0,Q-1]\!]}(\mathbf{x}+\alpha\mathbf{d}_k^{opt}+\mathbf{p})^t\mathbf{d}_k^p. \tag{9.12}$$

Le bruit de canal $\mathbf{p}$ n'étant pas connu lors de la procédure d'émission, le choix de $\mathbf{d}_k^{opt}$ est effectué dans un contexte non bruité ($\mathbf{p}=0$). Au regard de l'équation précédente, il conduit à un problème de maximisation conjointe : trouver le couple de vecteur $(\mathbf{d}_k^{opt},\mathbf{d}_k^q)$ tel que

$$(opt,q) = \arg\max_{(p,i)\in[\![0,Q-1]\!]^2}(\mathbf{x}+\alpha\mathbf{d}_k^p)^t\mathbf{d}_k^i. \tag{9.13}$$

Sachant le dictionnaire constitué de vecteurs orthogonaux ou bi-orthogonaux, cette relation se simplifie aisément sous la forme :

$$(opt,q) = \arg\max_{(p,i)\in[\![0,Q-1]\!]^2}\left(\mathbf{x}^t\mathbf{d}_k^i \pm \alpha N_s\delta(p,i)\right). \tag{9.14}$$

Il en est aisément déduit que $opt=q$: le signal $\mathbf{d}_k^{opt}$ choisi pour faciliter la détection de l'information k sera aussi celui détecté par le corrélateur $\mathbf{d}_k^q$ (si la détection sans bruit est correcte).

L'équation donnant le signal du sous-dictionnaire $\mathcal{D}_k$ le plus adapté à la transmission de l'information k peut finalement être réduite à la simple recherche du signal dont la corrélation avec le signal audio $\mathbf{x}$ est la plus grande :

$$opt = \arg\max_{q\in[\![0,Q-1]\!]}\mathbf{x}^t\mathbf{d}_k^q. \tag{9.15}$$

9.2.2 Performances théoriques dans le cas d'un canal non bruité

Cette section est dédiée à l'évaluation des performances théoriques de la chaîne de tatouage informée au sens de la probabilité d'erreur de transmission. Nous proposons de déterminer l'expression de la probabilité d'erreur dans le cas où le bruit de canal $\mathbf{p}$ est nul [2]. On supposera que les symboles sont émis de manière équiprobable.

9.2.2.1 Cas binaire

Dans un premier temps, nous nous intéressons plus particulièrement au cas d'une transmission binaire ($K=1$) : l'information à émettre k peut prendre seulement 2 valeurs $'0'$ ou $'1'$.

Ayant supposé l'équiprobabilité des bits émis et étant donnée la symétrie du problème, on peut supposer sans perte de généralité que l'information émise est $s='0'$. Or, pour transmettre le bit $s='0'$, on choisit parmi les vecteurs du sous-dictionnaire $\mathcal{D}_0$ le vecteur $\mathbf{d}_0^{opt}$ dont la

2. Les expressions de la probabilité d'erreur pourront aisément être généralisées au cas d'un bruit de canal non nul. Cette généralisation est justifiée lorsque le dictionnaire d'émission est de grande dimension : le système tend alors vers le schéma de Costa idéal dans lequel l'influence du signal audio sur les performances de transmission est complètement annulée ; seul le bruit de canal peut donc être source d'erreurs. Dans notre contexte, rappelons que le dictionnaire est de faible dimension. L'influence du signal audio reste ici présente et constitue une source d'interférences majeure dont nous souhaitons évaluer l'influence sur les performances de transmission.

corrélation avec le signal audio $\mathbf{x}$ est la plus grande. La probabilité d'erreur binaire devient alors la probabilité de recevoir l'information $'1'$ sachant que $\mathbf{d}_0^{opt}$ a été émis :

$$P_e = \text{Prob}\left(\hat{s} =' 1' | \mathbf{d}_0^{opt}\right). \tag{9.16}$$

La valeur de $\hat{s}$ est décidée en fonction du résultat des corrélations $\hat{\mathbf{c}} = \mathcal{D}^t(\mathbf{x} + \mathbf{t})$ entre le signal reçu et des vecteurs du dictionnaire. Sachant que $\mathbf{t} = \alpha \mathbf{d}_0^{opt}$, on peut donc représenter la métrique de corrélation $\hat{\mathbf{c}}$ dans le cas d'un dictionnaire orthogonal sous la forme normalisée suivante :

$$\tilde{\mathbf{c}} = \frac{1}{\alpha}\frac{1}{N_s}\hat{\mathbf{c}} = \begin{bmatrix} \tilde{c}_0^0 \\ \vdots \\ \tilde{c}_0^{opt} \\ \vdots \\ \tilde{c}_0^{Q-1} \\ \hdashline \tilde{c}_1^0 \\ \vdots \\ \tilde{c}_1^{Q-1} \end{bmatrix} = \begin{bmatrix} \tilde{x}_0^0 \\ \vdots \\ \tilde{x}_0^{opt} \\ \vdots \\ \tilde{x}_0^{Q-1} \\ \hdashline \tilde{x}_1^0 \\ \vdots \\ \tilde{x}_1^{Q-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ \hdashline 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{9.17}$$

où nous avons adopté les conventions de notations suivantes :

$$\tilde{\mathbf{c}} = [\tilde{c}_k^q]_{k \in [\![0,2^K-1]\!], q \in [\![0,Q-1[\![}, \quad \text{avec } \tilde{c}_k^q = \tfrac{1}{\alpha}\tfrac{1}{N_s}(\mathbf{x}+\mathbf{t})^t \mathbf{d}_k^q \tag{9.18}$$

$$\tilde{\mathbf{x}} = [\tilde{x}_k^q]_{k \in [\![0,2^K-1]\!], q \in [\![0,Q-1[\![}, \quad \text{avec } \tilde{x}_k^q = \tfrac{1}{\alpha}\tfrac{1}{N_s}\mathbf{x}^t \mathbf{d}_k^q. \tag{9.19}$$

Le vecteur de corrélation $\tilde{\mathbf{c}}$ peut donc être vu comme la réalisation de $2Q$ variables aléatoires, sommes d'une composante déterministe provenant de l'information émise et d'une composante aléatoire provenant du bruit audio $\mathbf{x}$.

Lors du procédé de détection, le vecteur du sous-dictionnaire $\mathcal{D}_0$ détecté est d_0^{opt}, puisqu'il est choisi parmi les vecteurs de $\mathcal{D}_0$ pour être celui de plus forte corrélation avec le bruit audio $\mathbf{x}$. Une erreur se produit s'il existe un vecteur du sous-dictionnaire $\mathcal{D}_1$ dont la corrélation avec le signal reçu est plus grande que $\hat{c}_0^{opt}$. La probabilité d'erreur peut donc être formalisée sous la forme :

$$P_e = \int_{-\infty}^{\infty} \text{Prob}(\text{erreur}|c) p_{\tilde{C}_0^{opt}}(c)\, dc = \int_{-\infty}^{\infty} (1 - P_c) p_{\tilde{C}_0^{opt}}(c)\, dc, \tag{9.20}$$

où $\text{Prob}(\text{erreur}|c)$ désigne la probabilité d'erreur sachant la valeur c de $\tilde{c}_0^{opt}$, P_c la probabilité d'une décision correcte sachant c et $p_{\tilde{C}_0^{opt}}$ est la densité de probabilité de la variable aléatoire $\tilde{C}_0^{opt}$ associée à $\tilde{c}_0^{opt}$.

Le calcul de la probabilité d'erreur peut donc être ramené aux calculs de la probabilité d'une décision correcte sachant la valeur de $\tilde{c}_0^{opt}$ et de la densité de probabilité de la variable aléatoire $\tilde{C}_0^{opt}$ associée à $\tilde{c}_0^{opt}$. Ces calculs dépendent de la configuration du dictionnaire et de l'espace signal qu'il engendre. Nous allons donc distinguer deux cas de figures : celui d'un dictionnaire orthogonal et celui d'un dictionnaire avec sous-dictionnaires biorthogonaux.

Dictionnaire orthogonal

Dans le cas d'un dictionnaire orthogonal, $\mathcal{D}$ engendre un espace de dimension $2Q$. De fait le

vecteur $\tilde{\mathbf{x}}$, puisque étant la projection du signal audio sur les vecteurs orthogonaux du dictionnaire, peut être vu comme la réalisation de $2Q$ variables aléatoires indépendantes, suivant la même loi normale de moyenne nulle et de variance $\tilde{\sigma}_x^2 = \dfrac{\sigma_x^2}{\alpha^2 N_s}$.

La probabilité d'une décision correcte connaissant la valeur de $\tilde{c}_0^{opt}$ est la probabilité que toutes les composantes $\tilde{c}_1^m$ de $\tilde{\mathbf{c}}$ relatives à la projection du signal reçu sur le sous-dictionnaire $\mathcal{D}_1$ soient inférieures à $\tilde{c}_0^{opt}$. Étant donnée l'indépendance des variables aléatoires auxquelles elles sont associées et la loi normale $\mathcal{N}(0, \dfrac{\sigma_x^2}{\alpha^2 N_s})$ qui décrit leurs propriétés statistiques, la probabilité d'une décision correcte peut s'écrire :

$$P_c = \text{Prob}\left(\forall q, \tilde{c}_1^q < \tilde{c}_0^{opt} | \tilde{c}_0^{opt}\right) = \left[\text{Prob}\left(\tilde{c}_1^0 < \tilde{c}_0^{opt} | \tilde{c}_0^{opt}\right)\right]^Q = \left[1 - \frac{1}{2}\text{erfc}\left(\sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}}\right)\right]^Q . \quad (9.21)$$

La densité de probabilité de la variable aléatoire $\tilde{C}_0^{opt}$ associée à $\tilde{c}_0^{opt}$ est dictée par le choix de l'indice *opt*. Rappelons que *opt* est par définition l'indice du vecteur du sous-dictionnaire $\mathcal{D}_0$ dont la corrélation avec le bruit audio est maximale. Ainsi :

$$\tilde{c}_0^{opt} = 1 + \tilde{x}_0^{opt}, \text{ avec } \tilde{x}_0^{opt} = \max_{q \in [\![0, Q-1]\!]} \tilde{x}_0^q. \quad (9.22)$$

Les $\tilde{x}_0^q$ étant issues de Q variables aléatoires indépendantes suivant la même loi normale $\mathcal{N}(0, \dfrac{\sigma_x^2}{\alpha^2 N_s})$, nous montrons en annexe que cette densité de probabilité s'écrit :

$$p_{\tilde{C}_0^{opt}}(c) = Q\sqrt{\frac{\alpha^2 N_s}{2\pi\sigma_x^2}}\left[1 - \frac{1}{2}\text{erfc}\left(\sqrt{\frac{\alpha^2 N_s}{\sigma_x^2}}\frac{(c-1)}{\sqrt{2}}\right)\right]^{Q-1} \exp^{-\frac{1}{2}\frac{\alpha^2 N_s}{\sigma_x^2}(c-1)^2} . \quad (9.23)$$

La probabilité d'erreur binaire dans le cas d'un dictionnaire orthogonal peut finalement être évaluée par :

$$P_e = \frac{Q}{\sqrt{\pi}}\int_{-\infty}^{\infty}\left[1 - \left(1 - \frac{1}{2}\text{erfc}(c)\right)^Q\right] \ldots$$
$$\ldots \left[1 - \frac{1}{2}\text{erfc}\left(c - \sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}}\right)\right]^{Q-1} \exp\left[\left(c - \sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}}\right)^2\right] dc. \quad (9.24)$$

Sous-dictionnaires biorthogonaux

On suppose maintenant que les sous-dictionnaires sont composés de vecteurs biorthogonaux engendrant deux espaces orthogonaux de dimensions $\dfrac{Q}{2}$. Cette fois, le vecteur $\tilde{\mathbf{x}}$ contient la réalisation de Q variables aléatoires indépendantes, suivant la même loi normale $\mathcal{N}(0, \dfrac{\sigma_x^2}{\alpha^2 N_s})$, et leurs opposées. Étant donnée la structure des sous-dictionnaires choisie section 9.2.1.2, les variables indépendantes sont celles associées à la projection du signal reçu sur les premiers vecteurs de chaque sous-dictionnaire, c'est à dire les $\tilde{c}_0^q$ et $\tilde{c}_1^q$ avec $q \in [\![0, \dfrac{Q}{2} - 1]\!]$.

La probabilité d'une décision correcte sachant $\tilde{c}_0^{opt}$ s'écrit maintenant :

$$P_c = \text{Prob}\left(\forall q \in [\![0, \frac{Q}{2} - 1]\!], |\tilde{c}_1^q| < \tilde{c}_0^{opt} | \tilde{c}_0^{opt}\right) = \prod_{q=0}^{\frac{Q}{2}-1} \text{Prob}\left(|\tilde{c}_1^q| < \tilde{c}_0^{opt} | \tilde{c}_0^{opt}\right). \tag{9.25}$$

On peut noter que $\tilde{c}_0^{opt}$ est nécessairement positif. En effet, l'indice *opt* est celui du vecteur $\mathbf{d}_0^{opt}$ de $\mathcal{D}_0$ dont la corrélation avec le signal audio est maximum. Si la corrélation était négative, le sous-dictionnaire contenant des vecteurs biorthogonaux, l'indice *opt* choisi serait alors celui du vecteur $-\mathbf{d}_0^{opt}$. P_c peut donc être exprimée sous la forme :

$$P_c = \sqrt{\frac{\sigma_x^2}{2\pi\alpha^2 N_s}} \int_{-\tilde{c}_0^{opt}}^{\tilde{c}_0^{opt}} exp\left(-\frac{\sigma_x^2}{2\alpha^2 N_s} c^2\right) dc = \text{erf}\left(\sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}} \tilde{c}_0^{opt}\right). \tag{9.26}$$

La densité de probabilité de la variable aléatoire $\tilde{C}_0^{opt}$ associée à $\tilde{c}_0^{opt}$ dépend cette fois de l'expression de $\tilde{c}_0^{opt}$:

$$\tilde{c}_0^{opt} = 1 + \tilde{x}_0^{opt}, \tilde{x}_0^{opt} = \max_{q \in [\![0, \frac{Q}{2}-1]\!]} |\tilde{x}_0^q|. \tag{9.27}$$

Nous montrons en annexe que l'expression de la densité de probabilité est la suivante :

$$p_{\tilde{C}_0^{opt}}(c) = \begin{cases} Q\sqrt{\frac{1}{2\pi}}\sqrt{\frac{\alpha^2 N_s}{\sigma_x^2}}\left[erf\left(\sqrt{\frac{\alpha^2 N_s}{\sigma_x^2}}\frac{(c-1)}{\sqrt{2}}\right)\right]^{\frac{Q}{2}-1} \exp^{-\frac{\alpha^2 N_s}{2\sigma_x^2}(c-1)^2} & \text{si } c > 1 \\ 0 & \text{si } c < 1 \end{cases}. \tag{9.28}$$

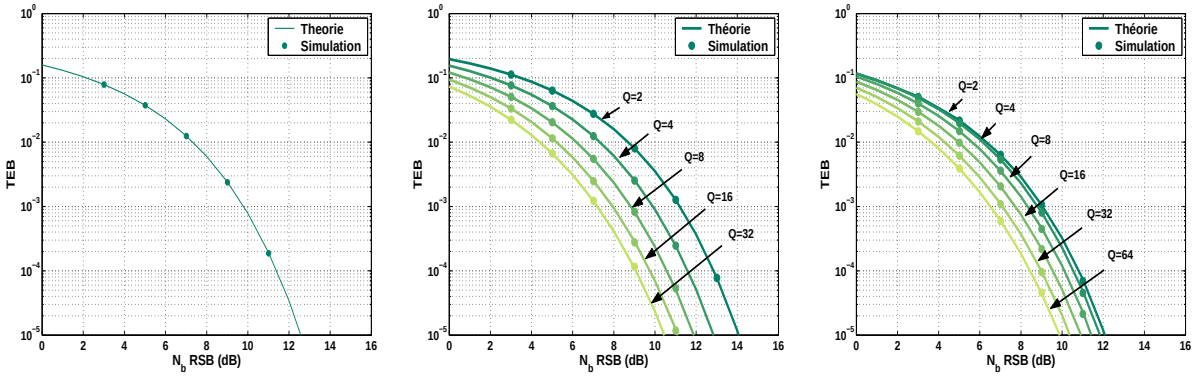
La probabilité d'erreur binaire peut finalement être évaluée par :

$$\begin{aligned} P_e &= \frac{Q}{\pi} \int_{\tilde{\sigma}_x}^{\infty} \left[1 - \text{erf}(c)^{\frac{Q}{2}}\right] \text{erf}\left(c - \sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}}\right)^{\frac{Q}{2}-1} \exp\left(-\left(c - \sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}}\right)^2\right) dc \\ &\text{avec } \tilde{\sigma}_x^2 = \frac{\alpha^2 N_s}{2\sigma_x^2}. \end{aligned} \tag{9.29}$$

Validation des expressions théoriques

Pour vérifier l'exactitude des expressions théoriques, nous avons simulé la transmission de 10 millions de bits à l'aide d'une procédure de Monte-Carlo pour différentes valeurs du paramètre de rapport signal à bruit $N_b\text{RSB} = \dfrac{\alpha^2 N_s}{\sigma_x^2}$ (défini précédemment) et différentes tailles des sous-dictionnaires Q. Cette configuration de simulation permet une fiabilité des mesures inférieures à 10^{-4} (dans le pire des cas). La figure 9.4 se propose de comparer les TEBs obtenus par simulations aux tracés de probabilité d'erreur théorique pour les deux configurations de dictionnaire envisagées. La figure 9.4 (a) rappelle les performances d'une modulation binaire antipodale non informée, les figures 9.4 (b) et (c) présente les résultats de la stratégie d'insertion proposée structurant le dictionnaire suivant le modèle de Costa dans le cas d'un dictionnaire orthogonal (figure (b)) et dans le cas d'un dictionnaire avec sous-dictionnaires biorthogonaux (figure (c)). Plusieurs conclusions peuvent être tirées de ces courbes :

- Les mesures de TEB obtenues par la simulation coïncident avec les tracés théoriques, validant ainsi les expressions théoriques de la probabilité d'erreur.



(a) Modulation binaire antipodale

(b) Avec dictionnaire orthogonal structuré

(c) Avec dictionnaire biorthogonal structuré

Figure 9.4 – Comparaison des TEBs obtenus par simulation et des tracés théoriques de la probabilité d'erreur dans différentes configurations : (a) une modulation non informée binaire antipodale, (b) une modulation informée basée sur un dictionnaire orthogonal structuré suivant le schéma de Costa, en fonction de la taille des sous-dictionnaires, (c) une modulation informée basée sur un dictionnaire biorthogonal structuré suivant le schéma de Costa, en fonction de la taille de ces sous-dictionnaires

- Quelle que soit la constellation engendrée par le dictionnaire, augmenter la taille des sous-dictionnaires Q et donc démultiplier les régions de détection d'une même information permet de diminuer la probabilité d'erreur de transmission et d'accroître les performances du système. L'apport d'un schéma informé suivant le modèle de Costa est donc mis une nouvelle fois en évidence.

- Les performances obtenues pour un dictionnaire avec sous-dictionnaires biorthogonaux sont globalement meilleures que celles obtenues avec un dictionnaire orthogonal. En effet, pour une même taille de l'espace signal, elles permettent d'obtenir un meilleur pavage de l'espace, c'est à dire de créer une meilleure alternance entre les régions de détection du bit $'0'$ et du bit $'1'$, comme cela avait été mis en évidence à la figure 9.3. Un dictionnaire avec sous-dictionnaires biorthogonaux augmente donc les chances de trouver une région de détection à proximité du signal audio.

9.2.2.2 Généralisation au cas d'une modulation 2^K-aire

La généralisation au cas d'une modulation 2^K-aire est quasiment immédiate. Il suffit d'augmenter le nombre de sous-dictionnaires et donc le nombre de composantes des vecteurs $\tilde{\mathbf{x}}$ et $\tilde{\mathbf{c}}$. La densité de probabilité de la variable aléatoire $\tilde{c}_0^{opt}$ reste inchangée. Par contre, la probabilité d'une décision correcte connaissant le vecteur $\mathbf{d}_0^{opt}$ émis doit maintenant prendre en compte toutes les composantes de $\tilde{\mathbf{c}}$ résultant de la projection du signal reçu sur les $2^K - 1$ sous-dictionnaires autres que $\mathcal{D}_0$. Les expressions de la probabilité d'erreur par symbole deviennent alors :

- dans le cas d'un dictionnaire orthogonal :

$$P_s = \frac{Q}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left[1 - \left(1 - \frac{1}{2}\mathrm{erfc}(c) \right)^{Q(2^K-1)} \right] \ldots$$
$$\ldots \left[1 - \frac{1}{2}\mathrm{erfc}\left(c - \sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}} \right) \right]^{Q-1} \exp\left(c - \sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}} \right)^2 dc. \quad (9.30)$$

- et dans le cas d'un dictionnaire avec sous-dictionnaire biorthogonaux :

$$P_s = \frac{Q}{\pi} \int_{\sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}}}^{\infty} \ldots \left[1 - \mathrm{erf}(c)^{\frac{Q}{2}(2^K-1)} \right] \ldots$$
$$\ldots \mathrm{erf}\left(c - \sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}} \right)^{\frac{Q}{2}-1} \exp\left(- \left(c - \sqrt{\frac{\alpha^2 N_s}{2\sigma_x^2}} \right)^2 \right) dc. \quad (9.31)$$

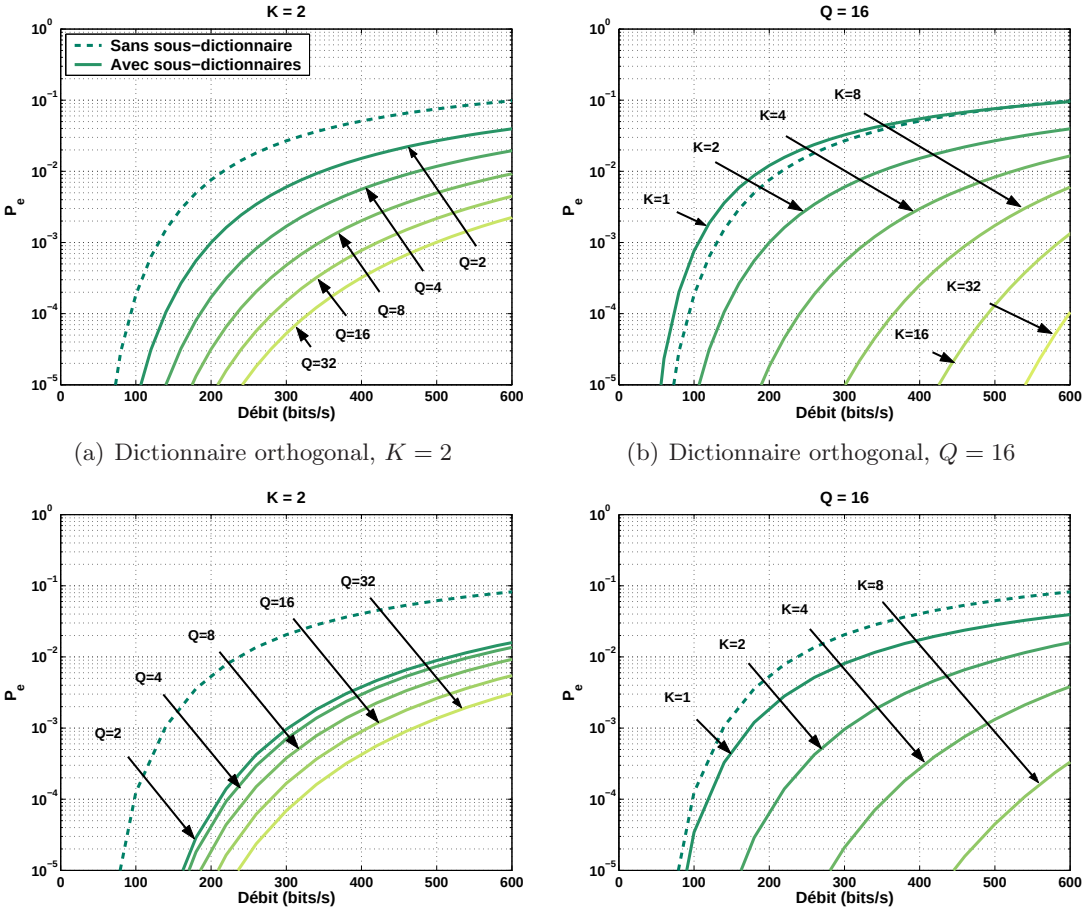
La probabilité d'erreur est ensuite déduite comme préalablement (c'est à dire à la section 4.4.1) de la probabilité d'erreur par symbole par la relation : $P_e = \frac{1}{2}\frac{2^K}{2^K-1}P_s$.

La figure 9.5 présente l'évolution de la probabilité d'erreur binaire en fonction du débit $R = \frac{KF_e}{N_s}$ lorsque le rapport signal à bruit est imposé à une valeur relativement peu favorable (ici -15 dB) pour différentes configurations et tailles de dictionnaire. Les figures 9.5 (a) et (b) s'intéressent au cas du dictionnaire orthogonal, tandis que les figures (c) et (d) présentent le cas d'un dictionnaire avec sous-dictionnaire biorthogonaux. L'influence de la taille des sous-dictionnaires Q (pour un paramètre d'encodage symbole K fixé) fait l'objet de figures (a) et (c). L'impact du paramètre K pour Q fixé est quant à lui mis en évidence par les figures (b) et (d). Trois remarques peuvent être faites au regard de ces courbes :

- Une stratégie basée sur le modèle de Costa et appliquée non plus à la transmission de bit mais à la transmission de symbole permet d'améliorer très nettement la fiabilité de transmission de système en jouant sur la taille du dictionnaire. Les probabilités d'erreur peuvent être diminuées quel que soit le débit aussi bien en augmentant le paramètre d'encodage K qu'en augmentant le nombre de vecteurs Q représentant chaque symbole.

- La prédominance de l'efficacité d'un dictionnaire avec sous-dictionnaires biorthogonaux par rapport à un dictionnaire orthogonal est également confirmée puisque les probabilités d'erreur pour une même taille de dictionnaire sont meilleures dans le premier cas que dans le second. Par la suite, nous ne nous intéresserons plus qu'au cas d'un dictionnaire avec sous-dictionnaires biorthogonaux.

- Pour finir, en comparant ces courbes aux performances des modulations M-aire pour la chaîne non informée, données figure 4.3 et 4.4 et rappelées ici pour $M = 2^K = 4$ par les tracés en pointillés, on met une nouvelle fois en évidence l'efficacité de la stratégie informée sur la stratégie non informée : pour un nombre de vecteurs du dictionnaire (avec $M = Q2^K$ dans le cas informé et $M = 2^K$ dans le cas non informé), les probabilités d'erreur sont nettement inférieures dans le cas du système informé que dans le cas du système non informé. Par exemple, pour un dictionnaire de 16 vecteurs, la probabilité d'erreur à 300 bit/s est de l'ordre de 8.10^{-3} pour le système informé (avec $K = 2$ et $Q = 4$) contre 8.10^{-2} pour le système non informé. Les TEBs pourront dans ce cas être divisés par 10 grâce à la stratégie d'insertion informée.



(a) Dictionnaire orthogonal, $K = 2$

(b) Dictionnaire orthogonal, $Q = 16$

(c) Sous-dictionnaires biorthogonaux, $K = 2$

(d) Sous-dictionnaires biorthogonaux, $Q = 16$

Figure 9.5 – Probabilité d'erreur binaire de la chaîne informée selon le modèle de dictionnaire de Costa pour un RSB de -15 dB en fonction du débit pour différentes configurations de dictionnaire : (a) un dictionnaire orthogonal avec K fixé à 2 et pour différentes valeurs de Q, (b) un dictionnaire orthogonal avec Q fixé à 16 et pour différentes valeurs de K, (c) un dictionnaire dont les sous-dictionnaires sont biorthogonaux avec K fixé à 2 et pour différentes valeurs de Q et (d) un dictionnaire dont les sous-dictionnaires sont biorthogonaux avec Q fixé à 16 et pour différentes valeurs de K. Les courbes en pointillés rappellent la probabilité d'erreur d'une modulation M-aire non informée (sans utilisation de sous-dictionnaire) lorsque le dictionnaire contient $M = 4$ vecteurs orthogonaux pour (a) et (b) et biorthogonaux pour (c) et (d).

9.2.3 Application au système de tatouage en boucle fermée

9.2.3.1 Principe

L'introduction d'une copie locale du récepteur à l'émetteur permet l'application quasi immédiate du dictionnaire structuré au système de tatouage en boucle fermée. Le dictionnaire d'émission $\mathcal{D}$ est modifié de sorte à contenir 2^K sous-dictionnaires $\{\mathcal{D}_k\}_{k\in[\![0,2^K-1]\!]}$ contenant chacun Q vecteurs biorthogonaux $\{\mathbf{d}_k^q\}_{q\in[\![0,Q-1]\!]}$ étalés dans la bande de fréquence $[0; F_c]$ kHz. La

copie locale permet d'estimer le dictionnaire de réception $\tilde{\mathcal{D}} = \{\tilde{\mathcal{D}}_k = \{\tilde{\mathbf{d}}_k^q\}_{q\in[\![0,Q-1]\!]}\}_{k\in[\![0,2^K-1]\!]}$ de même structure que le dictionnaire d'émission.

La sélection du vecteur $\mathbf{d}_k^{opt}$ du sous-dictionnaire $\mathcal{D}_k$ le plus adapté à la transmission de l'information k est liée cette fois à la corrélation entre la composante blanchie du signal audio $\tilde{\mathbf{x}}$ et le dictionnaire de réception $\tilde{\mathcal{D}}$ pour le récepteur par filtre blanchissant et $\mathcal{D}$ pour l'égaliseur de Larbi. Le choix de $\mathbf{d}_k^{opt}$, établi équation 9.15, devient :

- dans le cas du récepteur par filtre blanchissant,

$$opt = \arg \max_{q\in[\![0,Q-1]\!]} \tilde{\mathbf{x}}^t \tilde{\mathbf{d}}_k^q. \tag{9.32}$$

- dans le cas de l'égaliseur,

$$opt = \arg \max_{q\in[\![0,Q-1]\!]} \tilde{\mathbf{x}}^t \mathbf{d}_k^q. \tag{9.33}$$

Le facteur d'amplitude α n'intervenant pas dans le choix du vecteur du dictionnaire le plus adapté à la transmission de l'information k, il est donc aisé de combiner la stratégie d'insertion avec dictionnaire structuré au module de contrôle d'inaudibilité. Le signal modulé optimal est calculé dans un premier temps, puis le module de contrôle d'inaudibilité est mise en œuvre pour choisir le gain α adapté permettant le calcul du signal de tatouage $t(n)$.

Ce système sera par la suite référé sous le nom de **système de tatouage informé avec dictionnaire structuré et contrôle de l'inaudibilité**.

9.2.3.2 Spécificités de l'égaliseur de Wiener

Dans le cas du récepteur par filtre blanchissant, les filtres de la chaîne (le filtre de mise en forme perceptuelle et le filtre blanchissant) sont obtenus à partir du signal audio $\mathbf{x}$. Ils ne dépendent pas du vecteur du dictionnaire choisi pour la modulation de l'information k. Par contre dans le cas de l'égaliseur de Larbi, le filtre de Wiener est adapté au signal $\mathbf{z}$ somme de la composante audio blanchie par $\frac{1}{H(f)}$ et du signal modulé. Le filtre de Wiener est donc dépendant du vecteur du dictionnaire choisi pour la modulation et influence le signal audio $\tilde{\mathbf{x}}$ utilisé pour sélectionner ce vecteur.

La mise en œuvre nécessite donc de calculer Q filtres de Wiener, chacun adapté à l'insertion de l'un des Q vecteurs du sous-dictionnaire $\mathcal{D}_k$ susceptibles de moduler l'information k. De ces Q filtres sont déduites Q versions du signal audio $\tilde{\mathbf{x}}^q$ en amont du corrélateur, chaque version correspondant à l'insertion du vecteur $\mathbf{d}_k^q$. L'équation 9.33 établissant le choix du vecteur du dictionnaire le plus adapté à la transmission de l'information k devient :

$$opt = \arg \max_{q\in[\![0,Q-1]\!]} \tilde{\mathbf{x}}^{q\,t}\, \mathbf{d}_k^q. \tag{9.34}$$

9.2.3.3 Performances

Nous nous proposons de comparer les performances du système de tatouage en boucle fermée avec le dictionnaire structuré à celle du système de référence avec contrôle de l'inaudibilité établi au chapitre 5. Seuls deux critères de performances, l'inaudibilité et le TEB en l'absence de perturbations, sont considérés dans cette étude. La robustesse aux perturbations et temps de calcul seront étudiés sur le système en boucle fermée final. Ces deux critères de mesure sont donc évalués conformément au protocole expérimental établi dans la première partie. Le dictionnaire

Système	Système de référence	Système en boucle fermée	
Réception		Filtre blanchissant	Égaliseur de Larbi
Paramètres	$K = 4$, $Q = 1$, $M = 16$	$K = 4$, $Q = 4$, $M = 64$	
ODG	$-1,0$ $(0,07)$	$-1,02$ $(0,06)$	$-1,18$ $(0,10)$

Table 9.1 – Valeurs moyennes de l'ODG et écart-type pour le système de tatouage non informé de référence et les deux systèmes en boucle fermée, le premier basé sur le récepteur par filtrage blanchissant et le second sur l'égaliseur.

des deux systèmes de référence est constitué de 16 vecteurs biorthogonaux permettant l'insertion simultanée de $K = 4$ bits d'information sur chaque temps symbole. Le dictionnaire structuré du système informé est lui constitué de 64 vecteurs structuré en 16 sous-dictionnaires ($K = 4$) contenant chacun $Q = 4$ vecteurs biorthogonaux.

Inaudibilité

L'ODG moyenne obtenue après tatouage des 20 signaux du corpus à 206 bit/s par le système en boucle fermée est présentée table 9.1 pour les deux types de récepteurs et est comparée à celle du système de référence. Sa valeur est maintenue à proximité de -1 quel que soit le récepteur utilisé par la copie locale. Le tatouage obtenu, malgré la modification de l'étape de modulation, reste inaudible mais non gênant, ce que confirme des tests d'écoutes informels.

TEB pour un canal sans perturbation

Les figures 9.6 présentent les TEBs du système de tatouage en boucle fermée pour les deux récepteurs envisagés lorsque le canal ne subit pas de perturbations. Elles rappellent à titre de comparaison les performances du système non informé. Ces courbes confirment l'efficacité d'un dictionnaire structuré même de taille relativement faible. Les TEBs sont en effet au moins divisés par 2 par rapport au cas non informé. L'amélioration semble plus importante dans le cas de l'égaliseur par Wiener. En effet, ce schéma ne souffre pas ni de la perte de l'orthogonalité du dictionnaire de réception, contrairement au récepteur par filtre blanchissant, ni des erreurs d'estimation liés aux filtres utilisés dans la copie locale du récepteur à l'émetteur. Ainsi, grâce à cette stratégie, un TEB de 10^{-3} peut être obtenu pour un débit de l'ordre de 125 bit/s (lorsque le récepteur est le filtre blanchissant) voire 150 bit/s (lorsque le récepteur est l'égaliseur de Larbi) contre 80 bit/s pour un système non informé.

9.3 Tatouage de robustesse maximale aux perturbations additives

Dans cette deuxième section, nous développons une seconde famille de stratégies d'insertion ; leur objectif est de déterminer parmi l'ensemble des tatouages inaudibles et satisfaisant la détection correcte de l'information émise celui maximisant la robustesse de la transmission aux perturbations additives : cette robustesse peut s'exprimer soit sous la forme d'un paramètre de robustesse à maximiser soit sous la forme d'une expression de la probabilité d'erreur face à un BBAG à minimiser. Ces stratégies conduisent à deux problèmes d'optimisation que nous résolvons par trois algorithmes différents. Les performances de ces algorithmes sont analysées

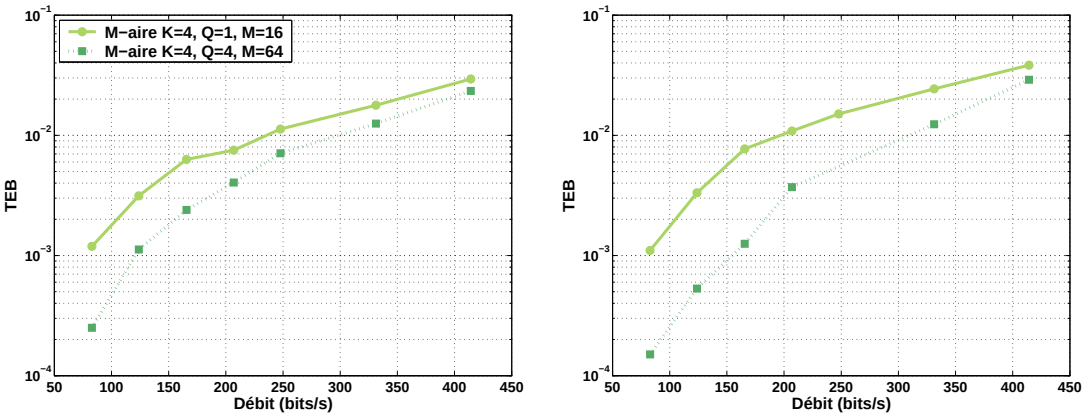


Figure 9.6 – Comparaison des performances du système non informé et de la stratégie d'insertion par dictionnaire structuré pour deux récepteurs : (a) le filtre blanchissant et (b) l'égaliseur de Larbi. Le canal ne subit pas de perturbations.

sur le modèle du système de tatouage avant d'être mises en œuvre sur les deux systèmes de tatouage en boucle fermée.

9.3.1 Vers la prise en compte de la perturbation

Reprenons le modèle de la chaîne de tatouage élaboré précédemment. Nous supposerons que la modulation est une modulation M-aire non informée utilisant un dictionnaire de M vecteurs orthogonaux (le dictionnaire n'est donc pas structuré comme il l'a été dans la section précédente).

Comme nous l'avons évoqué à la section 9.1, choisir pour signal de tatouage le vecteur du dictionnaire associé à l'information à transmettre (c'est à dire $\mathbf{d}_k$ si l'information est k) pondéré par le facteur d'amplitude α n'est pas la solution la plus adaptée : dans certaines configurations du signal audio, ce choix donne lieu à une erreur de transmission alors qu'il existe des signaux de tatouage autres que $\mathbf{d}_k$ permettant une transmission sans erreur. Nous nous proposons de définir une stratégie d'insertion informée qui permette de choisir parmi les signaux de tatouage possibles une unique solution. Cette solution doit satisfaire les deux contraintes du système de tatouage : elle doit garantir la transparence du tatouage et permettre une transmission sans erreur quelle que soit la perturbation effectuée sur le signal tatoué. La stratégie d'insertion se dirige naturellement vers le choix de l'unique tatouage maximisant la robustesse du système aux perturbations. La perturbation est modélisée par un bruit blanc additif Gaussien $\mathbf{p}$ suivant une loi normale $\mathcal{N}(0, \sigma_p^2)$. Contrairement au signal audio, supposé connu lors du procédé d'émission, cette perturbation sera supposée inconnue.

Le schéma du modèle informé est celui présenté figure 9.7. Cette modélisation permet à nouveau d'établir les conditions régissant le choix du tatouage $\mathbf{t}$ et du signal modulé $\mathbf{v} = \frac{1}{\alpha}\mathbf{t}$ dont il est issu. Ce tatouage doit être choisi dans l'espace signal à l'intersection de deux régions d'intérêt : la région d'inaudibilité et la région de détection robuste. L'espace signal étant l'espace engendré par les vecteurs du dictionnaire, le signal modulé $\mathbf{v}$ associé à $\mathbf{t}$ est recherché sous la

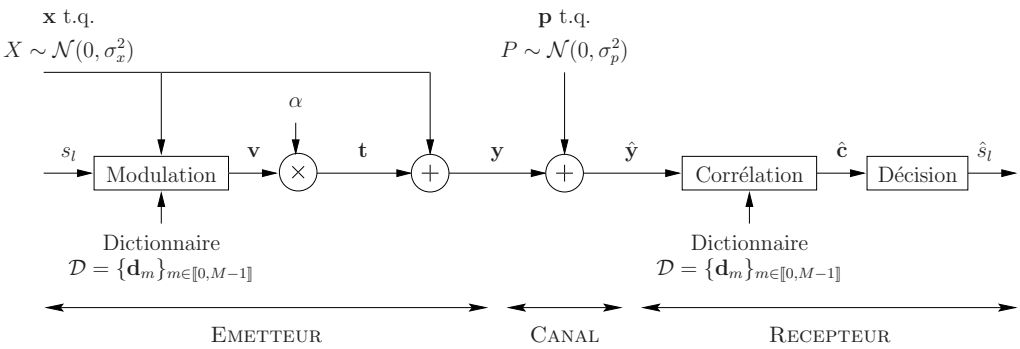


Figure 9.7 – Schéma du système de tatouage.

forme d'une combinaison linéaire de vecteurs du dictionnaire de coefficients $\{\lambda_m\}_{m \in [\![0, M-1]\!]}$:

$$\mathbf{v} = \sum_{m=0}^{M-1} \lambda_m \mathbf{d}_m = \mathcal{D}\lambda, \tag{9.35}$$

où λ est la représentation vectorielle des coefficients.

Rappelons l'expression des contraintes que doit satisfaire le tatouage $\mathbf{t}$ ou de façon équivalente le signal modulé $\mathbf{v}$:

- la contrainte d'inaudibilité est liée à la contrainte de puissance imposée au signal modulé $\mathbf{v}$. Le signal modulé conduira à un signal de tatouage inaudible si sa puissance est limitée à 1 :

$$\sigma_v^2 = \frac{1}{N_s}\mathbf{v}^t\mathbf{v} = \sum_{m=0}^{M-1} \lambda_m^2 = \lambda^t\lambda \leq 1. \tag{9.36}$$

  Cette limite traduit la puissance induisant la distorsion auditive maximum tolérée pour le signal de tatouage. Choisir un signal modulé de puissance inférieure à 1 peut donc être une solution acceptable en terme de dégradation auditive. En effet, pour un signal de puissance inférieure à 1, la dégradation auditive introduite par le tatouage est plus faible que la distorsion maximum tolérée. Néanmoins, elle conduira à un rapport signal à bruit plus faible et donc à une probabilité d'erreur de transmission plus grande.

- La contrainte de détection robuste s'exprime sous la forme de $M-1$ inégalités traduisant le résultat de la démodulation par corrélation :

$$\forall m \neq k, (\mathbf{x} + \alpha\mathbf{v} + \mathbf{p})^t \mathbf{d}_k > (\mathbf{x} + \alpha\mathbf{v} + \mathbf{p})^t \mathbf{d}_m. \tag{9.37}$$

La stratégie d'insertion informée se définit finalement comme le choix du signal modulé $\mathbf{v}$, combinaison linéaire des vecteurs du dictionnaire (9.35), qui satisfait à la fois l'équation (9.36) et le système d'inégalités (9.37) quel que soit le bruit de perturbation $\mathbf{p}$. Malheureusement, ce bruit n'est pas connu. Il faut donc substituer aux inégalités (9.37) un problème équivalent, prenant en compte un paramètre modélisant le bruit $\mathbf{p}$ et permettant de choisir plus aisément le signal modulé adapté. Deux approches peuvent être envisagées :

- une approche, que nous qualifierons de ***déterministe***, qui consiste à substituer le bruit $\mathbf{p}$ par un paramètre τ_p caractérisant la robustesse du système aux perturbations et que l'on souhaite rendre maximal.

- une approche, que nous qualifierons de ***probabiliste***. Le bruit $\mathbf{p}$ est supposé être la réalisation d'un vecteur aléatoire donc les composantes suivent une loi normale $\mathcal{N}(0, \sigma_p^2)$. La probabilité d'erreur peut alors être estimée et le signal modulé choisi de sorte à minimiser cette probabilité d'erreur pour une variance de bruit maximale.

Ces deux approches sont détaillées dans les sections suivantes.

9.3.2 Approche déterministe

9.3.2.1 Formalisation du problème

En regroupant les signaux connus à l'émetteur c'est à dire le signal audio $\mathbf{x}$ et de tatouage $\mathbf{t}$, le système d'inégalités traduisant la détection de l'information k peut être réécrit sous la forme :

$$\forall m \neq k, (\mathbf{x} + \alpha\mathbf{v})^t (\mathbf{d}_k - \mathbf{d}_m) > -\mathbf{p}^t (\mathbf{d}_k - \mathbf{d}_m) . \tag{9.38}$$

Ce système d'inégalités fait apparaître la projection du bruit $\mathbf{p}$ sur les directions $\mathbf{d}_k - \mathbf{d}_m$. Ces projections n'étant pas connues, un paramètre de robustesse τ_p peut leur être substitué, comme le suggère la stratégie ***Maximizing robustness*** proposée par Miller dans [MCB00]. Les conditions d'une détection correcte et robuste de l'information k s'écrivent finalement :

$$\forall m \neq k, (\mathbf{x} + \alpha\mathbf{v})^t (\mathbf{d}_k - \mathbf{d}_m) \geq \tau_p, \tag{9.39}$$

La détection sera d'autant plus robuste que τ_p est élevé. Or, maximiser τ_p revient à maximiser conjointement les projections $(\mathbf{x}+\alpha\mathbf{v})^t(\mathbf{d}_k - \mathbf{d}_m)$ donc en particulier la plus petite d'entre elles. Le signal modulé $\mathbf{v}$ doit donc être choisi de sorte que :

$$\mathbf{v} = \arg\max_{\mathbf{u}} \min_{m \neq k} (\mathbf{x} + \alpha\mathbf{u})^t (\mathbf{d}_k - \mathbf{d}_m). \tag{9.40}$$

La stratégie d'insertion informée, dans le cadre de l'approche déterministe, peut finalement être formalisée par un problème d'optimisation $(\mathcal{P}_1)$: trouver le signal modulé, combinaison linéaire des vecteurs du dictionnaire (9.35), dans la région d'inaudibilité définie par (9.36) et solution de l'équation de détection robuste (9.40) :

$$(\mathcal{P}_1) \ : \ \mathbf{v} \text{ tel que } \begin{cases} \mathbf{v} = \displaystyle\sum_{q=0}^{M-1} \lambda_q \mathbf{d}_q = \mathcal{D}\lambda \\ \sigma_v^2 = \dfrac{1}{N_s}\mathbf{v}^t\mathbf{v} = \displaystyle\sum_{q=0}^{M-1} \lambda_m^2 = 1 \\ \mathbf{v} = \arg\max_{\mathbf{u}} \min_{m \neq k} (\mathbf{x} + \alpha\mathbf{u})^t (\mathbf{d}_k - \mathbf{d}_m) \end{cases} . \tag{9.41}$$

9.3.2.2 Résolution du problème - Cas d'un dictionnaire de deux vecteurs

Dans le cas d'un dictionnaire contenant seulement deux vecteurs orthogonaux, le problème d'optimisation peut être résolu de manière formelle. L'équation de détection robuste peut en effet être simplifiée en supprimant l'étape de minimisation puisque le dictionnaire ne contient que deux vecteurs. Supposant par exemple que l'information à émettre est le bit '0' ($k = 0$), ces

deux vecteurs sont le vecteur $\mathbf{d}_k = \mathbf{d}_0$ associé à l'information que l'on souhaite émettre et le vecteur $\mathbf{d}_m = \mathbf{d}_1$ associé au bit '1'. L'équation de détection robuste devient :

$$\mathbf{v} = \arg\max_{\mathbf{u}} (\mathbf{x} + \alpha\mathbf{u})^t (\mathbf{d}_k - \mathbf{d}_m). \tag{9.42}$$

Quel que soit le bruit audio $\mathbf{x}$ et le facteur d'amplitude α, la projection du signal tatoué $\mathbf{x} + \alpha\mathbf{v}$ dans la direction $(\mathbf{d}_k - \mathbf{d}_m)$ est maximale si le signal modulé $\mathbf{v}$ est choisi de projection maximale sur $(\mathbf{d}_k - \mathbf{d}_m)$:

$$\mathbf{v} = \arg\max_{\mathbf{u}} \mathbf{u}^t (\mathbf{d}_k - \mathbf{d}_m). \tag{9.43}$$

Sa projection sur $(\mathbf{d}_k - \mathbf{d}_m)$ est maximale si $\mathbf{v}$ est choisi dans la direction de $(\mathbf{d}_k - \mathbf{d}_m)$, soit :

$$\mathbf{v} = \lambda (\mathbf{d}_k - \mathbf{d}_m). \tag{9.44}$$

$\mathbf{v}$ prend donc naturellement la forme d'une combinaison linéaire des deux vecteurs du dictionnaire de coefficient λ. Ce coefficient est choisi de sorte à garantir la contrainte d'inaudibilité (9.36) qui impose au signal modulé d'être de puissance unité. Ainsi :

$$\sigma_v^2 = \frac{1}{N_s}\mathbf{v}^t\mathbf{v} = 2\lambda^2 = 1 \Rightarrow \lambda = \frac{1}{\sqrt{2}}. \tag{9.45}$$

Contrairement à une stratégie d'insertion non informée, qui aurait choisi pour signal modulé $\mathbf{v} = \mathbf{d}_k$, le signal modulé optimal maximisant la robustesse du système aux perturbations est donc :

$$\mathbf{v} = \frac{1}{\sqrt{2}} (\mathbf{d}_k - \mathbf{d}_m). \tag{9.46}$$

La figure 9.8 met en évidence le signal modulé optimal pour une configuration de signal audio $\mathbf{x}$ favorable (c'est à dire garantissant l'existence d'une solution). La région de détection correcte est le demi-plan contenant l'ensemble des vecteurs dont la projection sur $\mathbf{d}_k - \mathbf{d}_m$ est positive. Cette région est le demi-plan supérieur à la droite portée par le vecteur $\mathbf{d}_k + \mathbf{d}_m$. En effet, cette droite est l'ensemble des points de corrélation avec $\mathbf{d}_k - \mathbf{d}_m$ nulle ; elle marque donc la frontière entre la région de détection de k et celle de m. La région de distorsion admissible est l'ensemble des signaux audio tatoués construits en additionnant le signal audio $\mathbf{x}$ aux signaux de tatouage $\mathbf{t} = \alpha\mathbf{v}$ de puissance α^2 : il s'agit donc d'un cercle centré sur $\mathbf{x}$ de rayon α. La projection de ce cercle sur la direction de détection $\mathbf{d}_k - \mathbf{d}_m$ permet de déterminer les valeurs limites prises par la fonctionnelle de l'équation de détection robuste $\mathbf{v} \mapsto (\mathbf{x} + \alpha\mathbf{v})^t (\mathbf{d}_k - \mathbf{d}_m)$. Ces limites sont obtenues en déterminant les droites parallèles à la frontière de détection qui tangentent le cercle de distorsion admissible. Le signal de tatouage optimal correspond donc à l'un de ces points de tangence. Il est porté par le vecteur orthogonal à $\mathbf{d}_k + \mathbf{d}_m$, qui n'est autre que la direction de détection $\mathbf{d}_k - \mathbf{d}_m$. Cette interprétation géométrique permet de retrouver l'expression du signal modulé optimal : $\mathbf{v} = \lambda(\mathbf{d}_k - \mathbf{d}_m)$. Elle met également en évidence une caractéristique intéressante du signal modulé optimal : ce signal est choisi dans une direction orthogonale à la frontière de détection, celle portée par le vecteur $\mathbf{d}_k - \mathbf{d}_m$.

Résoudre le problème d'optimisation dans le cas d'un dictionnaire contenant cette fois M vecteurs (avec $M > 2$) s'avère beaucoup plus délicat. Nous ne sommes pas parvenus à déterminer une solution formelle à ce problème. Deux algorithmes itératifs ont donc été envisagés pour permettre de déterminer le signal modulé adapté. Tous deux recherchent le signal modulé $\mathbf{v}$ sous la forme d'une combinaison linéaire des vecteurs du dictionnaire. Ils diffèrent dans l'approche envisagée pour solutionner le problème d'optimisation.

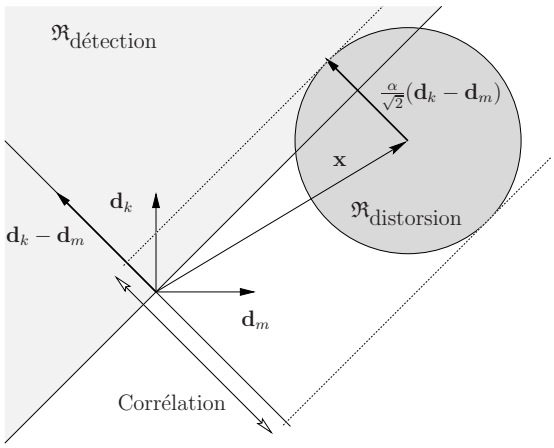


Figure 9.8 – Mise en évidence du signal modulé de robustesse maximale aux perturbations dans le cas d'un dictionnaire de deux vecteurs orthogonaux.

9.3.2.3 Algorithme Petit Pas

Cet algorithme, inspiré de celui proposé par Döerr dans [Doë01], construit itérativement le signal modulé optimal $\mathbf{v}$ en lui ajoutant des petites contributions l'amenant progressivement dans la direction optimale. Il se base principalement sur la résolution de l'équation de détection robuste 9.40 :

$$\mathbf{v} = \arg\max_{\mathbf{u}} \min_{m \neq k} (\mathbf{x} + \alpha \mathbf{u})^t (\mathbf{d}_k - \mathbf{d}_m). \tag{9.47}$$

Cette équation nous amène à définir la notion de ***vecteur le plus compétitif***. Ce vecteur, noté $\mathbf{d}_{comp}$, est le vecteur du dictionnaire $\mathcal{D}$ autre que $\mathbf{d}_k$ le plus susceptible d'être détecté et donc d'induire une erreur de transmission. Etant donné le récepteur par corrélation, il s'agit du vecteur du dictionnaire autre que $\mathbf{d}_k$ dont la corrélation avec le signal reçu $\mathbf{x} + \mathbf{t}$ (en l'absence de perturbation) est la plus grande :

$$comp = \arg\max_{m \neq k} (\mathbf{x} + \alpha \mathbf{v})^t \mathbf{d}_m. \tag{9.48}$$

Déterminer ce vecteur permet de se ramener temporairement à la configuration d'un dictionnaire de deux vecteurs orthogonaux : le vecteur $\mathbf{d}_k$ porteur de l'information à émettre k et le vecteur $\mathbf{d}_m$ le plus compétitif étant donné le signal de tatouage choisi. Dans cette configuration, nous avons montré précédemment que le signal modulé optimal devait être amené dans la direction orthogonale à la frontière de détection entre k et m, c'est à dire celle portée par le vecteur $\mathbf{d}_k - \mathbf{d}_m$. On peut donc construire le signal modulé optimal en lui ajoutant une contribution dans la direction $\mathbf{d}_k - \mathbf{d}_m$.

L'algorithme itératif, qui découle de cette constatation, est donné algorithme 9.1.

Trois remarques peuvent être faites à propos de cet algorithme :

- Dans le cas d'un dictionnaire de 2 vecteurs orthogonaux, cet algorithme permet de retrouver la solution optimale $\mathbf{v} = \dfrac{1}{\sqrt{2}}(\mathbf{d}_k - \mathbf{d}_m)$. En effet, le vecteur compétitif sélectionné par l'algorithme est toujours le vecteur $\mathbf{d}_m$ et chaque itération consiste à ajouter à $\mathbf{v}$ une fraction de $\mathbf{d}_k - \mathbf{d}_m$.

Étape 1 : Le signal modulé à l'itération 0, noté $\mathbf{v}^{(0)}$, est initialisé au vecteur nul :

$$\mathbf{v}^{(0)} = \sum_{m=0}^{M-1} \lambda_m^{(0)} \mathbf{d}_m \text{ avec } \forall m, \lambda_m^{(0)} = 0. \tag{9.49}$$

Étape 2 : Notons $\mathbf{v}^{(n)} = \sum_{m=0}^{M-1} \lambda_m^{(n)} \mathbf{d}_m$ le vecteur calculé à la n-ième itération. Le vecteur le plus compétitif $\mathbf{d}_{comp}$ étant donné le signal modulé $\mathbf{v}^{(n)}$ est déterminé en appliquant la relation (9.48) :

$$\text{comp} = \arg\max_{m \neq k} \left(\mathbf{x} + \alpha \mathbf{v}^{(n)} \right)^t \mathbf{d}_m. \tag{9.50}$$

Le signal modulé est modifié en le dirigeant dans le direction $\mathbf{d}_k - \mathbf{d}_{comp}$. Ainsi :

$$\mathbf{v}^{(n+1)} = \mathbf{v}^{(n)} + \rho(\mathbf{d}_k - \mathbf{d}_{comp}) \Leftrightarrow \left\{ \begin{array}{l} \lambda_k^{(n+1)} = \lambda_k^{(n)} + \rho \\ \lambda_{comp}^{(n+1)} = \lambda_{comp}^{(n)} - \rho \end{array} \right. , \tag{9.51}$$

où ρ est le pas de l'algorithme.
Si la puissance de $\mathbf{v}^{(n+1)}$ est inférieure à 1, l'étape 2 est réitérée. Sinon, le signal modulé optimal est obtenu en normalisant $\mathbf{v}^{(n+1)}$ en puissance :

$$\mathbf{v} = \frac{\mathbf{v}^{(n+1)}}{\sigma_{\mathbf{v}^{(n+1)}}} \text{ avec } \sigma^2_{\mathbf{v}^{(n+1)}} = \frac{\mathbf{v}^{(n+1)\,t} \mathbf{v}^{(n+1)}}{N_s}. \tag{9.52}$$

Algorithme 9.1: Algorithme Petit Pas

- Le critère d'arrêt choisi par l'algorithme ne conduit pas nécessairement au tatouage optimal. En effet, l'algorithme s'arrête dès que la puissance du signal modulé atteint 1 et non pas à l'unique point de tangence entre la région d'inaudibilité et la région de détection robuste.

- L'algorithme ne permet pas lors de sa mise en œuvre de déterminer s'il existe effectivement une solution garantissant la transmission robuste. En effet, le signal modulé est choisi de sorte à maximiser la métrique de corrélation à la réception que celle-ci soit positive ou non. Or si la valeur maximale de la corrélation est négative, une erreur de transmission est fortement prévisible : seul l'ajout d'un bruit en direction de $\mathbf{d}_k - \mathbf{d}_m$ pourrait ramener le signal reçu dans la région de détection de l'information à émettre k.

9.3.2.4 Algorithme d'Uzawa itéré

Cet algorithme résout le problème d'optimisation $(\mathcal{P}_1)$ en itérant la valeur du paramètre de robustesse τ_p jusqu'à atteindre sa valeur maximale admissible. Cette méthode peut être interprétée d'un point de vue géométrique : en reprenant l'exemple de la figure 9.1, il s'agit de translater la région de détection perpendiculairement à l'axe de détection $\mathbf{d}_k - \mathbf{d}_m$ jusqu'à atteindre la tangence entre la région de détection et la région de distorsion admissible. Le point de tangence marque le signal audio tatoué optimal. Cette interprétation est présentée figure 9.9,

où la région de détection devient d'autant plus grisée que τ_p est grand.

Plus concrètement, supposons pour l'instant τ_p fixé à une valeur arbitraire positive. Le signal modulé $\mathbf{v}$ doit être choisi de sorte à satisfaire la contrainte d'inaudibilité (9.36) et les contraintes de détection correcte et robuste (9.39) :

$$\begin{cases} \sigma_{\mathbf{v}}^2 = 1 \\ \forall m \neq k, (\mathbf{x} + \alpha\mathbf{v})^t(\mathbf{d}_k - \mathbf{d}_m) > \tau_p \end{cases} \tag{9.53}$$

Pour résoudre ce problème, nous choisissons d'imposer que $\mathbf{v}$ soit solution des contraintes de détection correcte et robuste. Plusieurs solutions existent à ce problème (en fait, l'ensemble de la région de détection est solution de ce problème). Pour ne sélectionner qu'un seul signal modulé $\mathbf{v}$, nous recherchons naturellement celui qui introduit le moins de dégradation auditive. La dégradation auditive étant liée à la puissance du signal modulé, nous sommes donc amenés à chercher le signal modulé $\mathbf{v}$ de puissance minimum sous la contrainte que ce signal soit solution des conditions de détection correcte et robuste. Cette approche conduit à la résolution du problème d'optimisation non linéaire sous contraintes, noté $(\mathcal{P}_2)$:

$$(\mathcal{P}_2) : \mathbf{v} = \arg\min_u \frac{1}{N_s}\mathbf{u}^t\mathbf{u} \text{ sous la contrainte } \mathcal{D}_d^t(\mathbf{x} + \alpha\mathbf{u}) \geq \tau_p, \tag{9.54}$$

où les contraintes de détection correcte et robuste sont présentées sous forme matricielle. Nous définissons donc $\mathcal{D}_d$ comme la matrice contenant les directions $(\mathbf{d}_k - \mathbf{d}_m)$, τ_p comme le vecteur colonne dont toutes les composantes sont égales à τ_p et nous généralisons le symbole $\geq$ au cas matriciel [3].

Résoudre un tel problème d'optimisation n'est pas immédiat du fait de la complexité des contraintes. La méthode d'Uzawa [Cia98] propose une alternative. Elle consiste à remplacer la minimisation sous contraintes par une suite de problèmes de minimisation sans contrainte plus simple à résoudre. Son principe est le suivant : notons $J_2(\mathbf{u}) = \frac{1}{N_s}\mathbf{u}^t\mathbf{u}$ le critère de puissance à minimiser. D'après les relations de Kuhn et Tucker [Cia98], J_2 atteint un minimum en $\mathbf{v}$ s'il existe un vecteur μ de composantes positives ou nulles, appelé ***multiplicateur de Lagrange généralisé***, tel que $(\mathbf{v}, \mu)$ est un point-selle de la fonction J_3 définie par :

$$J_3(\mathbf{v}, \mu) = J_2(\mathbf{v}) + \mu^t \left(\tau_p - \mathcal{D}_d^t(\mathbf{x} + \alpha\mathbf{v})\right). \tag{9.55}$$

Un point-selle est un point de la surface définie par $J_3(\mathbf{v}, \mu)$ qui minimise J_3 dans la direction $\mathbf{v}$ et maximise J_3 dans la direction de μ. Déterminer le multiplicateur de Lagrange μ pour lequel J_3 atteint son maximum permet donc de se ramener à un problème de minimisation sans contrainte.

Rechercher μ nécessite d'établir la relation liant $\mathbf{v}$ et μ. Cette relation est obtenue en annulant le gradient de J_3, noté $\nabla_{\mathbf{v}} J_3(\mathbf{v}, \mu)$, au point $\mathbf{v}$ où un minimum est atteint. Ainsi ,

$$\nabla_{\mathbf{v}} J_3(\mathbf{v}, \mu) = \frac{2}{N_s}\mathbf{v} - \mathcal{D}_d\,\mu = 0 \Rightarrow \mathbf{v} = \frac{N_s}{2}\mathcal{D}_d\,\mu. \tag{9.56}$$

En remplaçant $\mathbf{v}$ par son expression en fonction de μ, le problème d'optimisation $(\mathcal{P}_2)$ conduit au problème dual suivant :

$$\mu = \arg\max_\omega J_4(\omega) \text{ avec } J_4(\omega) = -\frac{N_s}{2}(\alpha - \frac{1}{2})\,\omega^t R_{\mathcal{D}_d}\omega + (\tau_p - \mathcal{D}_d^t\mathbf{x})^t\omega \text{ sous la contrainte } \mu \geq 0, \tag{9.57}$$

3. $\mathbf{a} \geq \mathbf{b}$ si et seulement si toutes les composantes du vecteur $\mathbf{a}$ sont supérieures ou égales aux composantes du vecteur $\mathbf{b}$.

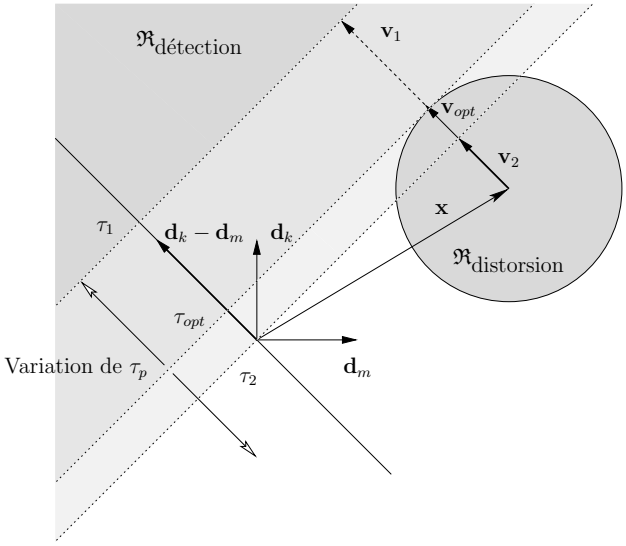


Figure 9.9 – Interprétation géométrique de l'algorithme d'Uzawa itéré.

où $R_{\mathcal{D}_d} = \mathcal{D}_d^t \mathcal{D}_d$ est la matrice d'autocorrélation de $\mathcal{D}_d$. Ce problème de maximisation est plus simple à résoudre que le problème $(\mathcal{P}_2)$ puisqu'il s'appuie sur une contrainte que l'on sait satisfaire. En effet, la méthode du gradient peut être utilisée avec l'opérateur de projection :

$$\mathrm{P} : \mu \rightarrow \begin{bmatrix} \max(\mu[0], 0) \\ \dots \\ \max(\mu[M-1], 0) \end{bmatrix} \tag{9.58}$$

La méthode d'Uzawa consiste finalement à faire converger μ et $\mathbf{v}$ conjointement et de manière itérative vers les valeurs optimales.

Supposons maintenant que nous connaissons la solution $\mathbf{v}$ obtenue pour la valeur de τ_p choisie arbitrairement. Trois cas de figure se présentent, comme mis en évidence figure 9.9 :

- la puissance de $\mathbf{v}$ est supérieure à 1, comme c'est le cas du signal $\mathbf{v}_1$ associé au paramètre τ_1 sur la figure 9.9. Pour satisfaire la contrainte de détection correcte et robuste, il faut donc choisir un signal modulé $\mathbf{v}$ introduisant une dégradation auditive supérieure à la limite admissible, symbolisée par la contrainte de puissance à 1 donnée par (9.36). Une robustesse τ_p ne peut être atteinte sans satisfaire la contrainte d'inaudibilité. τ_p doit donc être diminué.

- la puissance de $\mathbf{v}$ est inférieure à 1, comme pour le signal $\mathbf{v}_2$ associé à τ_2. Dans ce cas, la contrainte de détection correcte et robuste peut être satisfaite à l'aide d'un signal modulé introduisant une dégradation auditive plus faible que prévue. Il existe donc un signal modulé pour lequel le paramètre de robustesse est plus grand tout en garantissant une dégradation auditive limitée par la contrainte d'inaudibilité. τ_p peut donc être augmenté.

- la puissance de $\mathbf{v}$ est égale à 1, cas du signal $\mathbf{v}_{opt}$ associé à τ_{opt}. Le signal modulé optimal est trouvé ; le paramètre de robustesse τ_p a atteint sa valeur maximum admissible.

Ces cas de figure traduisent en réalité le fait que la puissance $\sigma_{\mathbf{v}}^2$ du signal modulé solution

du problème d'optimisation ($\mathcal{P}_2$) pour une certaine valeur de τ_p est une fonction croissante [4] de τ_p. Ainsi, faire converger la puissance du signal modulé vers sa valeur maximale admissible (c'est à dire 1) est équivalent à faire converger τ_p vers sa valeur maximale admissible (c'est à dire garantissant l'existence d'un signal modulé solution des conditions de détection correcte et robuste et de la contrainte d'inaudibilité).

L'algorithme proposé est donc un algorithme doublement itératif, puisqu'il consiste à itérer τ_p (de manière dichotomique), puis pour chaque valeur de τ_p à appliquer l'algorithme itératif d'Uzawa pour déterminer le signal modulé optimal associé à τ. Le détail de cet algorithme est présenté dans l'algorithme 9.2.

Le pas ρ, pour garantir la convergence de l'algorithme d'Uzawa, doit satisfaire la contrainte suivante : $0 < \rho < 2\dfrac{\beta}{||C||^2}$, où $\beta = \dfrac{2}{N_s}$ est un paramètre relatif à la fonctionnelle J_1 à minimiser et $C = \mathcal{D}_d^t\mathbf{x}$ caractérise les contraintes. L'objectif n'étant pas ici de démontrer l'efficacité de l'algorithme d'Uzawa, le lecteur se reportera à [Cia98] s'il souhaite obtenir une justification de ces paramètres. Le pas ρ que nous avons choisi est finalement :

$$\rho = \frac{4}{N_s||\mathcal{D}_d^t\mathbf{x}||^2}. \tag{9.62}$$

Deux remarques peuvent être faites à propos de cet algorithme :

- Son temps de calcul apparaît déjà être nettement plus élevé que celui de l'algorithme petit pas, étant donnée la double itération.

- Il offre néanmoins l'avantage de déterminer dès la construction du signal modulé optimal s'il existe effectivement une solution qui garantisse l'inaudibilité et les conditions de détection correcte de l'information. En effet, dans le cas d'une configuration de signal audio $\mathbf{x}$ fortement défavorable à la transmission de l'information k (pour laquelle la région de détection correcte et la région d'inaudibilité sont disjointes), l'algorithme converge vers une valeur de τ_p négative. Cela signifie qu'il existe un vecteur du dictionnaire, $\mathbf{d}_{comp}$, pour lequel quel que soit le signal modulé $\mathbf{v}$ de puissance 1 choisi :

  $$(\mathbf{x} + \alpha\mathbf{v})^t\mathbf{d}_k < (\mathbf{x} + \alpha\mathbf{v})^t\mathbf{d}_{opt}. \tag{9.63}$$

  A moins que le canal n'introduise un bruit $\mathbf{p}$ facilitant la transmission de k, une erreur de transmission est donc fortement prévisible.

9.3.3 Approche probabiliste

9.3.3.1 Principe

Reprenons le système d'inégalités traduisant les conditions de détection correcte de l'information k en présence d'un bruit de canal $\mathbf{p}$:

$$\forall m \neq k, (\mathbf{x} + \alpha\mathbf{v})^t(\mathbf{d}_k - \mathbf{d}_m) > -\mathbf{p}^t(\mathbf{d}_k - \mathbf{d}_m). \tag{9.64}$$

4. La monotonie de la fonction $f : \tau_p \mapsto \sigma_{\mathbf{v}}^2$ peut d'ailleurs être démontrée de manière formelle. En effet, notons $\mathbf{v}_1$ (respectivement $\mathbf{v}_2$) la solution du problème d'optimisation pour un paramètre de robustesse τ_1 (respectivement τ_2). Et supposons $\tau_1 < \tau_2$. $\mathbf{v}_2$ satisfait les contraintes de détection robuste pour τ_2, mais également pour τ_1 puisque : $\mathcal{D}_d^t(\mathbf{x} + \alpha\mathbf{v}_2) \geq \tau_2 > \tau_1$. Donc $\mathbf{v}_2$, au même titre que $\mathbf{v}_1$ est une solution potentielle du problème d'optimisation pour τ_1. On en déduit alors que : soit $\mathbf{v}_2 = \mathbf{v}_1$, soit $\mathbf{v}_2 \neq \mathbf{v}_1$ et dans ce cas, la puissance de $\mathbf{v}_2$ est nécessairement plus grande que celle de $\mathbf{v}_1$ puisque par définition, $\mathbf{v}_1$ est le vecteur de puissance la plus petite solution du problème d'optimisation. La fonction f est donc effectivement croissante.

Étape 0 : τ_p est initialisé à une valeur arbitraire τ_1 (ici choisie égale à la valeur maximale des composantes de $\mathcal{D}_d^t\mathbf{x}$).

Étape n : Étant donnée la valeur τ_n, l'algorithme d'Uzawa est appliqué pour résoudre le problème d'optimisation :

$$(\mathcal{P}_n) : \mathbf{v}_n = \arg\min_{\mathbf{u}} \frac{1}{N_s}\mathbf{u}^t\mathbf{u} \text{ sous la contrainte } \mathcal{D}_d^t(\mathbf{x}+\alpha\mathbf{v}) \geq \tau_n. \quad (9.59)$$

Il consiste en :

Sous-étape 0 : initialiser μ_0 à une valeur arbitraire.

Sous-étape 1 : appliquer la méthode du gradient pour rechercher le maximum de J_4 en définissant la suite μ_k par la relation de récurrence :

$$\begin{aligned} \mu_{k+1} &= \mathrm{P}\left(\mu_k - \rho\nabla_\mu J_4(\mu_k)\right), \text{ avec} \\ &\nabla_\mu J_4(\mu_k) = -N_s\left(\alpha - \frac{1}{2}\right)R_{\mathcal{D}_d}^t\mu_k + \tau_p - \mathcal{D}_d^t\mathbf{x}, \end{aligned} \quad (9.60)$$

où ρ est le pas de l'algorithme à définir.

Sous-étape 2 : définir parallèlement la suite $\mathbf{v}_k$ solution du problème d'optimisation $(\mathcal{P}_n)$ en utilisant la relation liant $\mathbf{v}_k$ à μ_k établie équation (9.56) :

$$\mathbf{v}_k = \frac{N_s}{2}\mathcal{D}_d\,\mu_k. \quad (9.61)$$

Sous-étape 3 : itérer le procédé jusqu'à ce que la variation de la puissance de $\mathbf{v}_k$ d'une itération à l'autre devienne négligeable (c'est à dire inférieur à un seuil τ_{var}).

Une fois la solution $\mathbf{v}_n$ obtenue, la puissance $\sigma^2_{\mathbf{v}_n}$ de $\mathbf{v}_n$ est comparée à la limite fixée par la contrainte d'inaudibilité, c'est à dire 1 :

- Si $\sigma^2_{\mathbf{v}_n}$ approche la valeur 1 à une erreur de τ_{var} près, l'algorithme s'arrête.
- Sinon, si $\sigma^2_{\mathbf{v}_n} > 1$ (respectivement $\sigma^2_{\mathbf{v}_n} < 1$), τ_n est décrémenté (respectivement incrémenté) en suivant une progression dichotomique[5].

Algorithme 9.2: Algorithme d'Uzawa itéré

Le bruit de canal $\mathbf{p}$ n'étant pas connu, nous choisissons de le modéliser comme la réalisation de N_s variables aléatoires de même loi normale $\mathcal{N}(0,\sigma_p^2)$. Cette modélisation permet d'établir l'expression de la probabilité d'erreur en fonction de $\mathbf{v}$, $\mathbf{x}$ et de la statistique de $\mathbf{p}$. Le signal modulé $\mathbf{v}$ pourra donc être choisi de sorte à maximiser la robustesse de la transmission, c'est à dire, le cas présent, de sorte à minimiser la probabilité d'erreur pour une variance de bruit σ_p^2 maximale.

Cette probabilité d'erreur dépend de la projection de $\mathbf{p}$ sur les directions de détection $\mathbf{d}_k - \mathbf{d}_m$. Pour faciliter la lecture des équations, nous définissons donc les $M-1$ vecteurs

$$\forall m \neq k, \mathbf{d}_{km} = \mathbf{d}_k - \mathbf{d}_m, \quad (9.65)$$

et les projections normalisées :

$$\tilde{x}_{km} = \frac{1}{\alpha}\frac{\mathbf{x}^t\mathbf{d}_{km}}{N_s},\ \tilde{v}_{km} = \frac{1}{\alpha}\frac{\mathbf{v}^t\mathbf{d}_{km}}{N_s},\ \text{ et } \tilde{p}_m = \frac{1}{\alpha}\frac{\mathbf{p}^t\mathbf{d}_m}{N_s}. \tag{9.66}$$

Les conditions de détection [6] peuvent être réécrites sous la forme :

$$\forall m \neq k, \tilde{x}_{km} + \tilde{v}_{km} > \tilde{p}_m - \tilde{p}_k. \tag{9.67}$$

$\mathbf{p}$ étant modélisé par N_s variables aléatoires $\mathcal{N}(0, \sigma_p^2)$, les M projections $\tilde{p}_m$ sont la réalisation de M variables aléatoires indépendantes de loi $\mathcal{N}(0, \tilde{\sigma}_p^2)$ avec $\tilde{\sigma}_p^2 = \dfrac{\sigma_p^2}{\alpha^2 N_s}$.

La probabilité d'une décision correcte quant au symbole k peut être évaluée en utilisant les conditions de transmission correcte :

$$P_c = \text{Prob}\left(\forall m \neq k, \tilde{x}_{km} + \tilde{v}_{km} > \tilde{p}_m - \tilde{p}_k\right). \tag{9.68}$$

Cette probabilité fait intervenir la composante aléatoire $\tilde{p}_k$ de loi $\mathcal{N}(0, \tilde{\sigma}_p^2)$, elle peut donc s'écrire sous la forme :

$$P_c = \int_{-\infty}^{\infty} \text{Prob}(\forall m \neq k, \tilde{x}_{km} + \tilde{v}_k m + \rho > \tilde{p}_m|\rho) p_{\tilde{P}_k}(\rho) d\rho, \tag{9.69}$$

où $\text{Prob}(\forall m \neq k, \tilde{x}_{km} + \tilde{v}_{km} + \rho > \tilde{p}_m|\rho)$ est la probabilité que les conditions de transmission robuste soient satisfaites sachant la valeur ρ de $\tilde{p}_k$ et $p_{\tilde{P}_k}(\rho) = \dfrac{1}{\sqrt{2\pi}\tilde{\sigma}_p} \exp\left(-\dfrac{1}{2}\dfrac{\rho^2}{\tilde{\sigma}_p^2}\right)$ est la densité de probabilité de la variable aléatoire $\tilde{P}_k$ associée à $\tilde{p}_k$. L'indépendance des variables aléatoires associées aux $M-1$ projections $\tilde{p}_m$ peut ensuite être mise à profiter pour simplifier le calcul de P_c en :

$$P_c = \int_{-\infty}^{\infty} \prod_{m \neq k} \text{Prob}(\tilde{x}_{km} + \tilde{v}_{km} + \rho > \tilde{p}_m|\rho) p_{\tilde{P}_k}(\rho) d\rho, \tag{9.70}$$

où $\text{Prob}(\tilde{x}_{km} + \tilde{v}_{km} + \rho > \tilde{p}_m|\rho) = 1 - \dfrac{1}{2}\text{erfc}\left(\dfrac{\tilde{x}_{km} + \tilde{v}_{km} + \rho}{\sqrt{2}\tilde{\sigma}_p}\right)$ est la probabilité que $\tilde{p}_m$ soit inférieur à $\tilde{x}_{km} + \tilde{v}_{km} + \rho$, connaissant la loi normale $\mathcal{N}(0, \tilde{\sigma}_p^2)$ qui régit la statistique de $\tilde{p}_m$.

Finalement la probabilité d'erreur par symbole est donnée par l'expression :

$$P_s = 1 - P_c = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_p} \int_{-\infty}^{\infty} \prod_{m \neq k} \left[1 - \frac{1}{2}\text{erfc}\left(\frac{\tilde{x}_{km} + \tilde{v}_{km} + \rho}{\sqrt{2}\tilde{\sigma}_p}\right)\right] \exp\left(-\frac{1}{2}\frac{\rho^2}{\tilde{\sigma}_p^2}\right) d\rho. \tag{9.71}$$

Cette approche consiste donc à trouver le signal modulé $\mathbf{v}$ satisfaisant à la contrainte d'inaudibilité qui minimise P_s pour une variance de bruit $\tilde{\sigma}_p^2$ maximale.

Ce problème de minimisation-maximisation conjointe ne peut être résolu, puisqu'il est défini à un degré de liberté près. En effet, pour une valeur des corrélations $\tilde{x}_{km} + \tilde{v}_{km}$ fixée (par exemple le tatouage optimal), la probabilité d'erreur P_s augmente avec la variance du bruit σ_p^2.

6. Nous retrouvons les conditions de détection du modèle réduit défini section **??**, mais appliquées cette fois non plus au dictionnaire $\mathcal{D}$ mais au dictionnaire contenant les directions de détection $\mathbf{d}_k - \mathbf{d}_m$, que nous avons appelé précédemment $\mathcal{D}_d$.

La résolution de ce problème nécessite donc soit de fixer la valeur de variance du bruit auquel le système doit être robuste, soit de fixer la probabilité d'erreur que l'on souhaite atteindre. Choisir la variance du bruit relève d'une stratégie d'insertion informée à robustesse constante, telle que l'a définie Miller [MCB00] ; nous avons volontairement mis de côté ce type de stratégie puisque le bruit de canal n'est pas supposé connu de l'émetteur. Nous préférons donc nous imposer une probabilité d'erreur à atteindre et maximiser la variance du bruit.

La stratégie d'insertion que nous proposons peut donc être formalisée de la façon suivante : trouver le tatouage $\mathbf{t}$ qui satisfait à la contrainte d'inaudibilité et permet d'atteindre une probabilité d'erreur P_s fixée à p pour une variance de bruit $\tilde{\sigma}_p^2$ maximale :

$$(\mathcal{P}_3) \quad : \quad \mathbf{v} \text{ tel que } \begin{cases} \sigma_v^2 = \frac{1}{N_s}\mathbf{v}^t\mathbf{v} \leq 1 \\ P_s = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_p}\int_{-\infty}^{\infty}\prod_{m\neq k}\left[1-\frac{1}{2}\text{erfc}\left(\frac{\tilde{x}_{km}+\tilde{v}_{km}+\rho}{\sqrt{2}\tilde{\sigma}_p}\right)\right]\exp\left(-\frac{1}{2}\frac{\rho^2}{\tilde{\sigma}_p^2}\right)d\rho = p \\ \tilde{\sigma}_p = \arg\max \sigma_p \end{cases} \tag{9.72}$$

9.3.3.2 Simplification du problème

L'expression de la probabilité d'erreur, relativement complexe et donc coûteuse à mettre en œuvre, peut être simplifiée : en effet, elle peut être sur-estimée par l'expression suivante :

$$P_s \leq P_s' = 1 - \prod_{m\neq k}\left[1-\frac{1}{2}\text{erfc}\left(\frac{\tilde{x}_{km}+\tilde{v}_{km}}{2\tilde{\sigma}_p}\right)\right] \tag{9.73}$$

D'après l'équation 9.68, la probabilité de détection sans erreur s'écrit :

$$P_c = 1 - P_e = \text{Prob}\left(\forall m \neq k, \tilde{x}_{km} + \tilde{v}_{km} > \tilde{p}_m - \tilde{p}_k\right). \tag{9.74}$$

Introduisons les $M-1$ variables $\tilde{p}_{mk} = \tilde{p}_m - \tilde{p}_k = -\dfrac{1}{\alpha}\dfrac{\mathbf{p}^t\mathbf{d}_{km}}{N_s}$. Ces variables sont de moyenne nulle $E[\tilde{p}_{mk}] = 0$, de variance $E[\tilde{p}_{mk}^2] = 2\tilde{\sigma}_p^2$ et de coefficient de corrélation pour chaque paire de variables :

$$\forall m \neq q \neq k,\ \rho_{mq} = \frac{E[\tilde{p}_{mk}\tilde{p}_{qk}]}{2\sigma_p^2} = \frac{1}{2}\mathbf{d}_{km}^t\mathbf{d}_{kq}. \tag{9.75}$$

Si le coefficient de corrélation n'est pas négatif, nous en déduisons que :

$$\text{Prob}\left(\tilde{x}_{km}+\tilde{v}_{km} > \tilde{p}_{mk}, \tilde{x}_{kq}+\tilde{v}_{kq} > \tilde{p}_{qk}\right) \geq \text{Prob}\left(\tilde{x}_{km}+\tilde{v}_{km} > \tilde{p}_{mk}\right)\text{Prob}\left(\tilde{x}_{kq}+\tilde{v}_{kq} > \tilde{p}_{qk}\right). \tag{9.76}$$

Pour appréhender la démonstration de cette inégalité, prenons l'exemple de deux variables aléatoires Gaussiennes $\tilde{p}_1$ et $\tilde{p}_2$ de moyenne nulle et de variance unité. Dans le cas particulier où $\tilde{x}_{km} + \tilde{v}_{km} = 0$, nous déduisons que $\text{Prob}\left(0 > \tilde{p}_1, 0 > \tilde{p}_2\right) = \dfrac{1}{4} + \dfrac{\arcsin(\rho)}{2\pi}$, où ρ est le coefficient de corrélation. Ainsi, lorsque $\rho > 0$, l'inégalité :

$$\text{Prob}\left(0 > \tilde{p}_1, 0 > \tilde{p}_2\right) > \frac{1}{4} = \text{Prob}\left(0 > \tilde{p}_1\right)\text{Prob}\left(0 > \tilde{p}_2\right)$$

est effectivement obtenue. Elle peut être généralisée au cas de $M-1$ variables Gaussiennes, ce qui permet d'obtenir l'inégalité (9.74) sous la forme :

$$1 - P_e \geq 1 - P_e' = \prod_{m\neq k}\text{Prob}\left(\tilde{x}_{km}+\tilde{v}_{km} > \tilde{p}_{mk}\right), \tag{9.77}$$

à la condition que les variables $\tilde{p}_{mk}$ soient de corrélations non nulles. Dans le cas du modèle du système de tatouage dont le dictionnaire est biorthogonal, $\rho_{mq} = \frac{1}{2}\mathbf{d}_{km}^t\mathbf{d}_{kq} \geq 0$; en effet $\rho_{mq} = 1$ si $\mathbf{d}_m$ est orthogonal à $\mathbf{d}_q$ et ρ_{mq} si $\mathbf{d}_m = -\mathbf{d}_q$. Dans le cas du système de tatouage en boucle fermée, les $\{\mathbf{d}_m\}$ correspondent aux vecteurs du dictionnaire de réception ; si le récepteur est basé sur le filtre blanchissant, l'orthogonalité du dictionnaire peut être perdue suite aux opérations de filtrages par $H(f)$ et $G(f)$. Néanmoins des simulations montrent que dans tous les cas, les coefficients de corrélations sont positifs.

Finalement, $\tilde{p}_{mk}$ suivant une loi normale de moyenne nulle et de variance $2\sigma_p^2$, $\text{Prob}\,(\tilde{x}_{km} + \tilde{v}_{km} > \tilde{p}_{mk}) = 1 - \frac{1}{2}\text{erfc}\left(\frac{\tilde{x}_{km} + \tilde{v}_{km}}{2\tilde{\sigma}_p}\right)$, conduisant à l'inégalité 9.73.

Au regard de cette nouvelle estimation de la probabilité d'erreur, le problème d'optimisation peut être réécrit sous la forme : trouver

$$(\mathcal{P}_3') : \mathbf{v} \text{ tel que } \begin{cases} \sigma_v^2 = \frac{1}{N_s}\mathbf{v}^t\mathbf{v} \leq 1 \\ P_s' = 1 - \prod_{m\neq k}\left(1 - \frac{1}{2}\text{erfc}\left(\frac{\tilde{x}_{km}+\tilde{v}_{km}}{2\tilde{\sigma}_p}\right)\right) = p \\ \tilde{\sigma}_p = \arg\max \sigma_p \end{cases} \tag{9.78}$$

9.3.3.3 Résolution du problème

Le problème d'optimisation $(\mathcal{P}_3')$ auquel nous avons à faire face est beaucoup plus complexe que les problèmes précédents. En effet les contraintes (principalement la probabilité d'erreur fixée) sont fortement non linéaires. La littérature [Fle87] propose plusieurs algorithmes d'optimisation puissants permettant de résoudre les problèmes d'optimisation où la fonction à minimiser est non linéaire et les contraintes quadratiques. Nous nous sommes plus particulièrement intéressés à la méthode de Lagrange-Newton également appelé Sequential Quadratique Programming (SQP), puisqu'une implémentation très pratique est proposée par la toolbox d'optimisation de Matlab©. Cette méthode permet de résoudre une minimisation sous contraintes non linéaires de manière itérative en se basant à chaque itération sur des approximations linéaires des contraintes.

L'algorithme de mise en œuvre proposé est un algorithme itératif similaire à l'algorithme d'Uzawa itéré précédemment. Il consiste à itérer la variance du bruit $\tilde{\sigma}_p$ jusqu'à atteindre sa valeur maximale admissible. Pour chaque valeur de $\tilde{\sigma}_p$ l'algorithme SQP est mis en œuvre pour déterminer le signal modulé $\mathbf{v}(\tilde{\sigma}_p)$ de puissance minimale qui satisfait à la contrainte de probabilité d'erreur fixée :

$$\begin{aligned} (\mathcal{P}_4) \quad : \quad & \mathbf{v}(\tilde{\sigma}_p) = \arg\min_u \frac{1}{N_s}\mathbf{u}^t\mathbf{u} \\ & \text{sous la contrainte } P_s' = 1 - \prod_{m\neq k}\left(1 - \frac{1}{2}\text{erfc}\left(\frac{\tilde{x}_{km} + \tilde{v}_{km}}{2\tilde{\sigma}_p}\right)\right) = p. \end{aligned} \tag{9.79}$$

La solution optimale est obtenue lorsque la puissance de $\mathbf{v}(\tilde{\sigma}_p)$ converge vers 1. Le détail de l'algorithme est présenté dans l'algorithme 9.3.

Étape 0 : $\tilde{\sigma}_p$ est initialisé à une valeur arbitraire $\tilde{\sigma}_1$.

Étape n : Étant donné la valeur $\tilde{\sigma}_n$ de $\tilde{\sigma}_p$, l'algorithme d'optimisation SQP est appliqué pour résoudre le problème d'optimisation :

$$(\mathcal{P}_n) \quad : \quad \mathbf{v}_n = \arg\min_u \frac{1}{N_s}\mathbf{u}^t\mathbf{u}$$
$$\text{avec} \begin{cases} 1 - \prod_{m\neq k}\left(1 - \frac{1}{2}\text{erfc}\left(\frac{\tilde{x}_{km}+\tilde{v}_{km}}{2\tilde{\sigma}_n}\right)\right) = p, \\ \tilde{v}_{km} = \frac{1}{\alpha}\frac{\mathbf{v}^t(\mathbf{d}_k-\mathbf{d}_m)}{N_s} \end{cases} \tag{9.80}$$

Une fois la solution $\mathbf{v}_n$ obtenue, la puissance $\sigma^2_{\mathbf{v}_n}$ de $\mathbf{v}_n$ est comparée à la limite fixée par la contrainte d'inaudibilité, c'est à dire 1 :
· Si $\sigma^2_{\mathbf{v}_n}$ approche la valeur 1 à une erreur de τ_{var} près, l'algorithme s'arrête.
· Sinon, si $\sigma^2_{\mathbf{v}_n} > 1$ (respectivement $\sigma^2_{\mathbf{v}_n} < 1$), $\tilde{\sigma}_n$ est décrémenté (respectivement incrémenté) en suivant une progression dichotomique.

Algorithme 9.3: Algorithme itéré basé sur une SQP

9.3.4 Performances des trois stratégies pour le modèle de tatouage

9.3.4.1 Comparaison des trois stratégies

Cinq critères de comparaison sont mis en œuvre pour évaluer et comparer l'efficacité de ces trois stratégies :

- la mesure du **TEB**,

- un **critère de robustesse**, remplaçant la mesure relativement coûteuse du TEB pour différents bruits de canal $\mathbf{p}$ (et qui s'avérera non significative au regard des TEBs obtenus). La détection étant liée à la corrélation entre le signal reçu et les vecteurs du dictionnaire, ce critère de robustesse consiste à évaluer l'écart entre la valeur de la corrélation pour le vecteur $\mathbf{d}_k$ à recevoir et celle du vecteur autre que $\mathbf{d}_k$ le plus compétitif. Ce critère peut être formalisé par :
$$CR = \frac{1}{N_s}(\mathbf{x}+\mathbf{t})^t\mathbf{d}_k - \max_{m\neq k}\frac{1}{N_s}(\mathbf{x}+\mathbf{t})^t\mathbf{d}_m, \tag{9.81}$$
où $\mathbf{t}$ est le tatouage solution du problème d'optimisation. Plus cet écart est élevé, plus la détection est susceptible d'être robuste à l'ajout d'un bruit additif.

- un **critère de puissance**. Contrairement au cas d'un système non informé où le tatouage est $\mathbf{t} = \mathbf{d}_k$, le système informé fournit : $\mathbf{t} = \lambda_k\mathbf{d}_k + \sum \lambda_{m\neq k}\mathbf{d}_m$. Si un tel tatouage permet de se prémunir contre la détection des vecteurs compétiteurs lors de la détection, il le fait au détriment de l'énergie qu'il consacre à émettre $\mathbf{d}_k$. Ce critère se propose donc d'évaluer l'énergie consacrée à $\mathbf{d}_k$. Il est défini par :
$$CP = \left(\frac{\mathbf{t}^t\mathbf{d}_k}{N_s}\right)^2 \approx \lambda_k, \tag{9.82}$$
où $\mathbf{t}$ est le tatouage solution du problème d'optimisation.

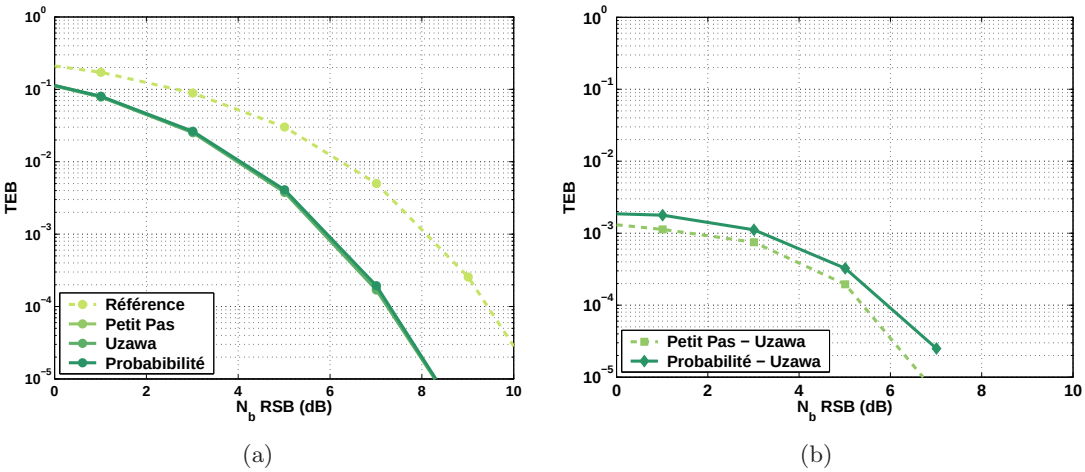


Figure 9.10 – (a) TEBs en fonction du paramètre N_bRSB obtenus par le modèle non informé et les 3 stratégies d'insertion informées basées respectivement sur l'algorithme Petit Pas, l'algorithme d'Uzawa itéré et l'algorithme probabiliste. (b) l'algorithme d'Uzawa étant le plus efficace, différence entre les TEBs obtenus par Petit Pas et l'algorithme d'Uzawa ainsi que celle entre les TEBs de l'algorithme probabiliste et l'algorithme d'Uzawa en fonction du paramètre N_bRSB.

Plus la valeur de CP est grande et plus la stratégie a tendance à sous-estimer l'influence des vecteurs compétiteurs. Inversement, plus CP est petit, plus l'énergie consacrée à se prémunir contre les vecteurs compétiteurs est grande, mais plus la transmission de $\mathbf{d}_k$, porteur de l'information à recevoir, est négligée.

- un **critère de rapport signal à bruit**. Dans un contexte de stratégie non informée, l'ensemble des composantes du signal audio est considéré comme un bruit. Pour les stratégies informées, nous avons établi que sa composante dans la direction $\mathbf{d}_k$ peut aider à la transmission de k. Dans cette direction, elle contribue à la transmission du message et peut être vue comme une partie du signal utile, porteur de l'information k à transmettre. Ses composantes dans les directions $\mathbf{d}_m$ autres que $\mathbf{d}_m$ nuisent à la détection de k et sont vues comme des composantes de bruits. Le tatouage optimal est choisi pour deux raisons : diriger le signal transmis dans la direction de $\mathbf{d}_k$ pour transmettre k, mais également tenter d'annuler les composantes du bruit $\mathbf{x}$ dans les directions $\mathbf{d}_m$ pour réduire le bruit audio. Un critère de rapport signal à bruit peut donc s'avérer significatif. Ce rapport est défini par :

$$RSB = \frac{((\mathbf{x}+\mathbf{t})^t\,\mathbf{d}_k)^2}{\sum_{m\neq k}((\mathbf{x}+\mathbf{t})^t\,\mathbf{d}_m)^2} \tag{9.83}$$

- un **critère de coût en temps de calcul** pour évaluer la rapidité de convergence des algorithmes et de leur mise en œuvre.

La figure 9.10 présente les TEBs binaires obtenus pour les trois stratégies d'insertion informées en fonction du paramètre N_bRSB déjà utilisé précédemment. Le protocole expérimental est le suivant : le modèle du système de tatouage est simulé par une séquence de 100000 bits (conduisant à une fiabilité de transmission de l'ordre de 10^{-3} dans le pire des cas) en utilisant un

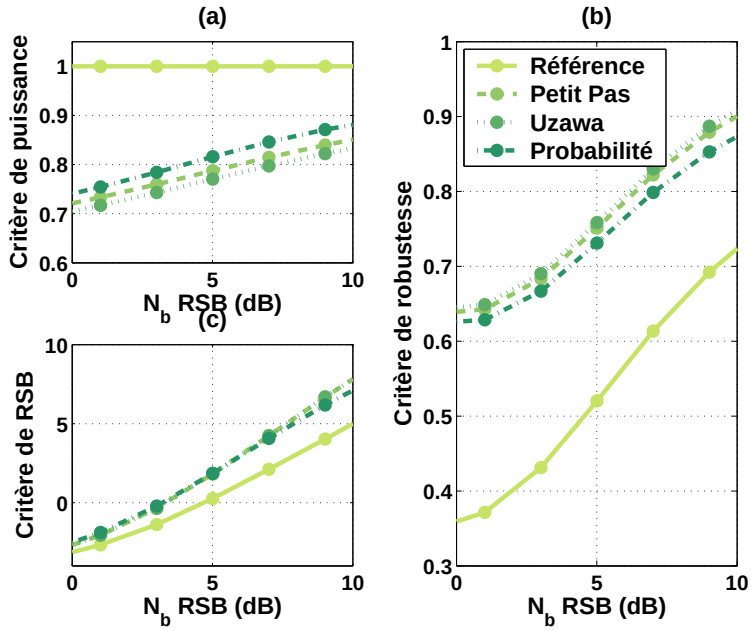


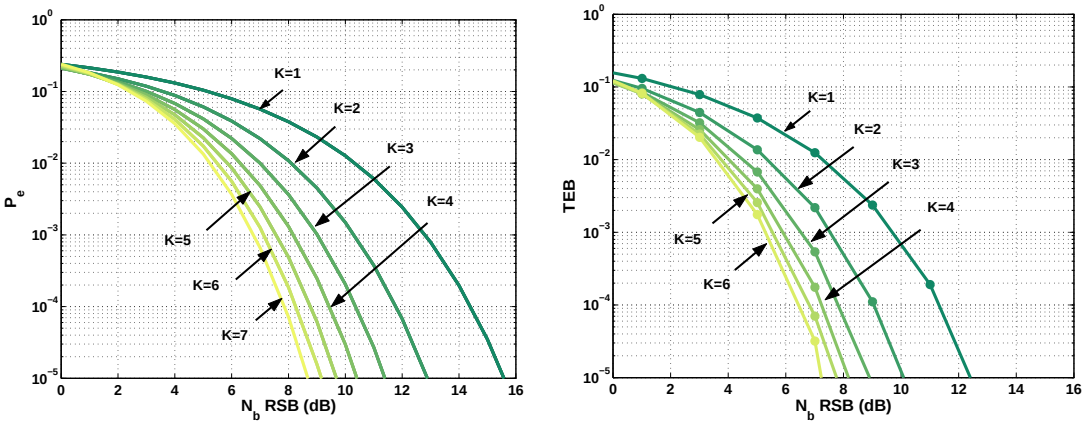
Figure 9.11 – Critères de puissance (a), de robustesse (b) et de RSB (c) en fonction du paramètre N_bRSB pour les trois stratégies d'insertion informée et la stratégie non informée.

dictionnaire de $M = 16$ vecteurs permettant l'encodage simultané de $K = 4$ bits. Les paramètres de mises en œuvre des stratégies d'insertion sont les suivants : la méthode Petit Pas utilise un pas $\rho = 0,02$ et la probabilité d'erreur face au bruit de canal à atteindre pour la méthode probabiliste est $p = 10^{-3}$. Cette figure présente également les performances de la stratégie d'insertion non informée. Ces courbes mettent en évidence l'apport des stratégies d'insertion informées que nous avons proposées sur les TEBs du système : un gain croissant lorsque la puissance du bruit audio diminue est obtenu allant d'un rapport 2 à 0 dB jusqu'à un rapport 100 à 10 dB. Par contre, les TEBs obtenus par les trois stratégies sont quasiment identiques ; la stratégie basée sur l'algorithme d'Uzawa se détache légèrement des deux autres en atteignant des TEBs légèrement plus faibles. Nous présentons donc sur cette figure la visualisation des deux différences entre les TEBs obtenus par la méthode Petit Pas et la méthode probabiliste et les TEBs d'Uzawa. Cette différence est effectivement très faible et de l'ordre de grandeur de la fiabilité des mesures. Du point de vue des TEBs, les trois méthodes sont donc équivalentes.

La figure 9.11 présente les critères de robustesse, de puissance et de rapport signal à bruit obtenus par les trois stratégies d'insertion informée et la stratégie non informée en fonction du paramètre N_bRSB. En terme de puissance, la figure 9.11 (a) montre que la puissance consacrée à la transmission de l'information k est plus grande dans le cas de l'algorithme probabiliste et plus faible dans le cas de l'algorithme d'Uzawa. En matière de robustesse, les conclusions sont inversées : la figure 9.11 (b) montre qu'Uzawa révèle de la plus grande valeur du critère de robustesse et ce quel que soit le paramètre N_bRSB tandis que la méthode probabiliste est la moins robuste. Cette figure met également en évidence le gain apporté par la stratégie d'insertion informée sur la robustesse du système par rapport à celui des stratégies non informées : l'écart de corrélations pour le vecteur que l'on souhaite recevoir et le vecteur compétitif le plus susceptible d'être détecté est deux fois plus grand dans le cas informé que dans le cas non informé. Cet

Approche	Petit Pas	Uzawa	Probabilité
Rapport de coût	51	1635	8638

Table 9.2 – Coût en temps de calcul des trois stratégies d'insertion informée sous la forme du rapport entre le temps nécessaire au calcul du tatouage informé et le temps pour le système non informé. Les simulations sont effectuées sous Matlab©.



(a) (b)

Figure 9.12 – TEBs en fonction du paramètre N_bRSB et de la taille M du dictionnaire pour (a) le système non informé et (b) le système informé basé sur l'algorithme Petit Pas.

apport est également mis en évidence par le critère RSB, présenté figure 9.11 (c). Le RSB est plus favorable à la transmission de l'information k pour les stratégies d'insertion informée que pour la stratégie non informé. Ce critère ne permet malheureusement pas d'établir de distinction entre les 3 stratégies d'insertion informées. De cette comparaison, nous conclurons donc qu'Uzawa semble être la stratégie la plus robuste, suivi par l'algorithme Petit Pas puis la méthode probabiliste.

Les trois méthodes informées étant quasi équivalentes en terme de TEB et de robustesse, le critère de coût en temps de calcul devient prépondérant pour sélectionner la méthode la plus efficace. La table 9.2 présente le rapport entre le temps en seconde nécessaire au calcul du tatouage optimal sous Matlab©et le temps de calcul du tatouage non informé pour les trois algorithmes. L'algorithme d'Uzawa comme l'algorithme probabiliste mettant en œuvre une double itération sont très coûteux temps de calcul contrairement à l'algorithme Petit Pas dont le coût reste raisonnable.

Au regard des coûts et des performances en terme de TEB, nous ne conserverons par la suite que l'algorithme Petit Pas, qui constitue finalement le meilleur compromis entre TEB, robustesse et coût des trois stratégies d'insertion.

9.3.4.2 Performances de l'algorithme Petit Pas

Dans cette section, nous proposons de comparer les performances en terme de TEB du système non informé et du système informé (avec l'algorithme Petit Pas) en fonction de la taille du dictionnaire. La figure 9.12 (a) rappelle les performances du système non informé en terme de

TEB en fonction du paramètre N_bRSB. La figure 9.12 (b) présente les TEBs du système informé lorsque la chaîne simule le tatouage de 10^6 bits. Ces courbes mettent en évidence une nouvelle fois le gain apporté par la stratégie d'insertion informée : en effet, en introduisant la stratégie d'insertion informée et sans modifier la taille du dictionnaire, les TEBs peuvent être divisés par 2 pour des configurations peu favorables du signal audio (conduisant à un paramètre N_bRSB de 0 dB) et être divisés par 100 pour des configurations très favorables (où N_bRSB est de 10 dB).

9.3.5 Application au système de tatouage en boucle fermée

L'introduction des trois stratégies d'insertion informée proposant de maximiser la robustesse du tatouage aux perturbations sous contrainte d'inaudibilité a jusqu'à maintenant été réalisée sur le modèle présenté figure 8.3. Des trois stratégies proposées nous n'avons conservé que la stratégie basée sur l'algorithme Petit Pas qui offre le meilleur compromis entre TEB et coût en temps de calcul. Il s'agit maintenant d'étudier les performances de cette stratégie sur le système réel donné figure 8.1 et 8.2.

9.3.5.1 Recherche du tatouage optimal pour un dictionnaire non structuré

L'objectif de la stratégie d'insertion appliquée à la copie locale de l'émetteur est de déterminer le signal de tatouage $\mathbf{t}$ en construisant le signal modulé $\mathbf{v}$ dont il est issu et sa version filtrée $\tilde{\mathbf{v}}$ qui satisfait la contrainte d'inaudibilité et de détection correcte et robuste dépendant du récepteur choisi. Il s'agit :

- dans le cas du récepteur par filtre blanchissant, de trouver $\mathbf{v}$ et $\tilde{\mathbf{v}}$ tel que :

$$\begin{cases} \sigma_v^2 = \frac{\mathbf{v}^t\mathbf{v}}{N_s} \leq 1 \\ \forall m \neq k, (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \tilde{\mathbf{d}}_k \geq (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \tilde{\mathbf{d}}_m. \end{cases} \tag{9.84}$$

- dans le cas de l'égaliseur de Wiener, de trouver $\mathbf{v}$ et $\tilde{\mathbf{v}}$ tel que :

$$\begin{cases} \sigma_v^2 = \frac{\mathbf{v}^t\mathbf{v}}{N_s} \leq 1 \\ \forall m \neq k, (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \mathbf{d}_k \geq (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \mathbf{d}_m. \end{cases} \tag{9.85}$$

La recherche du signal modulé sous la forme d'une combinaison linéaire des vecteurs du dictionnaire d'émission permet une recherche aisée des signaux $\mathbf{v}$ et $\tilde{\mathbf{v}}$. En effet, la linéarité des opérations de filtrage conjointement au fait que les filtres ne modifient pas la dimension de l'espace signal entre l'émission et la réception établit que :

$$\mathbf{v} = \mathcal{D}\lambda \Leftrightarrow \tilde{\mathbf{v}} = \tilde{\mathcal{D}}\lambda \tag{9.86}$$

Il est donc possible de rechercher de manière conjointe le signal modulé dans l'espace d'insertion et sa version filtrée dans l'espace de détection.

Le formalisme unifié de la copie locale du récepteur à l'émetteur nous a permis de résoudre un problème similaire : trouver $\mathbf{v}$ tel que

$$\begin{cases} \sigma_v^2 = \frac{\mathbf{v}^t\mathbf{v}}{N_s} \leq= 1 \\ \forall m \neq k, \forall m \neq k, (\mathbf{x} + \alpha\mathbf{v} + \mathbf{p})^t \mathbf{d}_k \geq (\mathbf{x} + \alpha\mathbf{v} + \mathbf{p})^t \mathbf{d}_m. \end{cases} \tag{9.87}$$

En introduisant un paramètre de robustesse au bruit, ce problème devient équivalent à un problème d'optimisation sous contrainte, que nous avons choisi de résoudre par l'algorithme itératif Petit Pas :

$$\mathbf{v} = \arg\max_{\mathbf{u}} \min_{m \neq k} (\mathbf{x} + \alpha \mathbf{u})^t (\mathbf{d}_k - \mathbf{d}_m) \text{ sous la contrainte } \frac{\mathbf{u}^t \mathbf{u}}{N_s} = 1. \tag{9.88}$$

L'algorithme Petit Pas est un algorithme itératif qui construit le signal modulé en lui ajoutant des composantes dans la direction $\mathbf{d}_k - \mathbf{d}_m$ où $\mathbf{d}_m$ est le vecteur le plus compétitif à chaque itération (c'est à dire minimisant la corrélation $(\mathbf{x} + \alpha \mathbf{v})^t (\mathbf{d}_k - \mathbf{d}_m)$ dont dépend la détection de k).

La mise en œuvre pour la copie locale du récepteur à l'émetteur est donc immédiate, sachant la relation entre $\tilde{\mathbf{v}}$ et $\mathbf{v}$. Elle s'attache au problème d'optimisation :

- dans le cas du récepteur par filtre blanchissant,

$$\begin{aligned} \tilde{\mathbf{v}} &= \arg\max_{\tilde{\mathbf{u}}} \min_{m \neq k} (\tilde{\mathbf{x}} + \alpha \tilde{\mathbf{u}})^t \left(\tilde{\mathbf{d}}_k - \tilde{\mathbf{d}}_m\right) \\ &\quad \text{sous la contrainte } \frac{\mathbf{u}^t \mathbf{u}}{N_s} = 1 \text{ avec } \mathbf{u} = \mathcal{D}\lambda \text{ et } \tilde{\mathbf{u}} = \tilde{\mathcal{D}}\lambda \end{aligned} \tag{9.89}$$

- dans le cas de l'égaliseur,

$$\begin{aligned} \tilde{\mathbf{v}} &= \arg\max_{\tilde{\mathbf{u}}} \min_{m \neq k} (\tilde{\mathbf{x}} + \alpha \tilde{\mathbf{u}})^t (\mathbf{d}_k - \mathbf{d}_m) \\ &\quad \text{sous la contrainte } \frac{\mathbf{u}^t \mathbf{u}}{N_s} = 1 \text{ avec } \mathbf{u} = \mathcal{D}\lambda \text{ et } \tilde{\mathbf{u}} = \tilde{\mathcal{D}}\lambda \end{aligned} \tag{9.90}$$

Les modifications de l'algorithme Petit Pas en sont aisément déduites. A chaque itération, le signal modulé $\mathbf{v}$ et sa version filtrée $\tilde{\mathbf{v}}$ sont construites conjointement en ajoutant au signal modulé $\mathbf{v}$ de petites contributions dans la direction $\mathbf{d}_k - \mathbf{d}_m$, ce qui est équivalent à ajouter à $\tilde{\mathbf{v}}$ une contribution dans la direction $\tilde{\mathbf{d}}_k - \tilde{\mathbf{d}}_m$. L'indice m est choisi à chaque itération pour être l'indice du vecteur le plus compétitif à la réception. Ce vecteur est :

- le vecteur $\tilde{\mathbf{d}}_m$ minimisant la corrélation $(\tilde{\mathbf{x}} + \alpha \tilde{\mathbf{v}})^t \left(\tilde{\mathbf{d}}_k - \tilde{\mathbf{d}}_m\right)$, dans le cas du filtre blanchissant,
- le vecteur $\mathbf{d}_m$ minimisant la corrélation $(\tilde{\mathbf{x}} + \alpha \tilde{\mathbf{v}})^t (\mathbf{d}_k - \mathbf{d}_m)$, dans le cas de l'égaliseur de Wiener.

9.3.5.2 Prise en compte du module de contrôle d'inaudibilité

Contrairement aux cas de figures précédents, le tatouage $\mathbf{t}$ issu de la stratégie d'insertion informée est dépendant du facteur d'amplitude α choisi spécifiquement pour maintenir une distorsion acceptable entre le signal audio original et sa version tatouée. Inversement, la dépendance du gain α vis à vis du tatouage $\mathbf{t}$ est peu accentuée, comme nous l'avons évoqué section 6.4.3.

La stratégie d'insertion informée est donc mise en œuvre deux fois sur chaque fenêtrage du signal audio de durée N_m sur lequel est calculé α. La première permet de construire le signal de tatouage $\mathbf{t}$ adapté pour la valeur de α obtenue à l'issu du tatouage de la fenêtre précédente. Le module de contrôle est ensuite mis en œuvre pour calculer la valeur adaptée de α sur la fenêtre en cours. Puis le tatouage optimal est calculé en applicant une seconde fois la stratégie d'insertion informée au regard de la valeur adaptée de α.

9.3.5.3 Spécificités de l'égaliseur de Wiener

Dans le cas du récepteur par filtre blanchissant, les filtres de la chaîne (le filtre de mise en forme perceptuelle et le filtre blanchissant) ne dépendent pas du choix du signal modulé $\mathbf{v}$. Dans le cas de l'égaliseur de Larbi, le filtre de Wiener est adapté au signal $\mathbf{z}$ somme de la composante audio blanchie par $\frac{1}{H(f)}$ et du signal modulé. Le filtre de Wiener est donc dépendant du signal modulé et influence le calcul du signal audio blanchi $\tilde{\mathbf{x}}$ utilisé dans l'algorithme de sélection du signal modulé $\mathbf{v}$ adapté.

L'analyse du critère de puissance donné figure 9.11 (a) montre que la majeure partie de la puissance du signal modulé est consacrée à la transmission de $\mathbf{d}_k$ (plus de 70% dans les configurations les plus défavorables). Le signal modulé adapté $\mathbf{v}$ est donc peu différent de $\mathbf{d}_k$. Or, ce dernier correspond au signal modulé construit à partir de la stratégie d'insertion non informée. Le filtre de Wiener peut donc en première approximation être calculé pour le signal modulé $\mathbf{d}_k$.

9.3.5.4 Performances

Nous nous proposons une nouvelle fois de comparer les performances du système de tatouage en boucle fermée avec le dictionnaire structuré à celle des deux systèmes de tatouage établis précédemment : le système de référence non informé avec contrôle de l'inaudibilité établi au chapitre 5 et le système de tatouage informé avec dictionnaire structuré proposé section 9.2. Seuls deux critères de performances, l'inaudibilité et le TEB en l'absence de perturbations, sont considérés dans cette étude. La robustesse aux perturbations et temps de calcul seront étudiés sur le système en boucle fermée final. Ces deux critères de mesure sont évalués conformément au protocole expérimental établi dans la première partie. Les caractéristiques des dictionnaires utilisées sont les suivantes : le dictionnaire pour le système de référence et le système informé basé sur l'algorithme Petit Pas est constitué de 16 vecteurs biorthogonaux permettant l'insertion simultanée de $K = 4$ bits d'information sur chaque temps symbole ; le dictionnaire structuré du système informé basé sur le modèle de Costa est lui constitué de 64 vecteurs structuré en 16 sous-dictionnaires ($K = 4$) contenant chacun $Q = 4$ vecteurs biorthogonaux.

Inaudibilité

Le respect de la contrainte d'inaudibilité est évalué par la mesure de l'ODG moyenne (et de l'écart type), obtenue par tatouage des 20 signaux du corpus à 206 bit/s. La table 9.3 présente les résultats obtenue pour la stratégie d'insertion informée utilisant l'algorithme Petit Pas en fonction du récepteur considéré. Cette table rappelle également les performances du système de référence non informé et du système informé basé sur un dictionnaire structuré suivant le modèle de Costa. On constate que l'introduction de la stratégie d'insertion informée ne modifie pas les performances du système en matière d'inaudibilité. L'ODG obtenue est toujours environ de -1 avec un écart-type entre les valeurs relativement faible. Le respect de la contrainte d'inaudibilité par l'algorithme Petit Pas est donc efficace.

TEB en l'absence de perturbations de canal

Les figures 9.13 présentent les TEBs du système de tatouage en boucle fermée lorsque la stratégie d'insertion est basée sur l'algorithme Petit Pas en fonction du débit et pour les deux récepteurs envisagés. Le canal est supposé ne pas subir de perturbations. Elles rappellent à titre

Système	Non informé de référence	Informé	
		avec dictionnaire structuré	avec l' algorithme Petit Pas
Filtre blanchissant	$-1,0$ $(0,07)$	$-1,02$ $(0,06)$	$-1,01$ $(0,06)$
Égaliseur de Larbi	$-1,0$ $(0,07)$	$-1,18$ $(0,10)$	$-1,01$ $(0,06)$

Table 9.3 – Valeurs moyennes de l'ODG et écart-type pour le système de tatouage non informé de référence et les deux systèmes en boucle fermée, le premier basé sur le récepteur par filtrage blanchissant et le second sur l'égaliseur de Larbi.

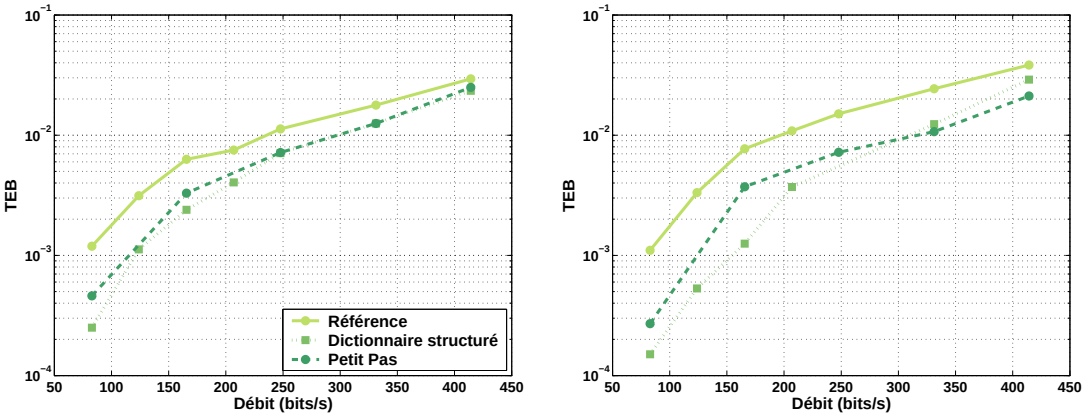


Figure 9.13 – Comparaison des TEBs du système avec stratégie d'insertion informée basée sur l'algorithme Petit Pas aux TEBs du système non informé et de la stratégie d'insertion par dictionnaire structuré pour deux récepteurs : (a) le filtre blanchissant et (b) l'égaliseur de Larbi. Le canal ne subit pas de perturbations.

de comparaison les performances du système non informé et celle du système informé avec un dictionnaire structuré.

Ces courbes confirment l'apport de la stratégie d'insertion informée basée sur l'algorithme Petit Pas sur la stratégie non informée : pour un même dictionnaire, choisir pour tatouage non plus le vecteur du dictionnaire porteur de l'information mais une combinaison linéaire de ces vecteurs maximisant la robustesse aux bruits de canal sous contrainte d'inaudibilité permet de diminuer les TEBs. En effet, quel que soit le débit et quel que soit le récepteur envisagé, les TEBs avec la stratégie d'insertion informée sont au moins divisés par 2 en comparaison du système non informé. Une transmission de TEB 10^{-3} peut grâce à la stratégie d'insertion informée être obtenue à un débit de 125 bit/s contre 80 dans le cas non informé.

Ces courbes montrent également que les performances de la stratégie informée avec dictionnaire structuré sont globalement meilleures que celle de la stratégie avec l'algorithme Petit Pas. Cette conclusion est d'autant plus marquée pour des débits inférieurs à 250 bit/s et lorsque le récepteur est l'égaliseur de Larbi.

Pour conclure, les deux stratégies d'insertion informée permettant d'améliorer chacune à leur façon les performances du système, il paraît alors judicieux de combiner ces deux stratégies.

Cette combinaison établira le système de tatouage en boucle fermée final. Elle fait l'objet de la dernière section.

9.4 Système de tatouage en boucle fermée avec dictionnaire structuré et choix d'un tatouage de robustesse maximale aux perturbations

Dans cette section, nous proposons de construire un système de tatouage en boucle fermée qui combine les deux stratégies d'insertion informée établies précédemment. Ce système pourra de fait bénéficier de l'apport du dictionnaire structuré suivant le modèle de Costa et de la procédure choisissant un tatouage de robustesse maximale aux perturbations.

9.4.1 Recherche du tatouage optimal pour un dictionnaire structuré

Nous supposons maintenant que le dictionnaire est structuré sous la forme de sous-diction naires $\{\mathcal{D}_m\}_{m\in[\![0,M-1]\!]}$, chaque sous-dictionnaire $\mathcal{D}_m$ contenant Q vecteurs $\{\mathbf{d}_m^q\}_{q\in[\![0,Q-1]\!]}$ tous susceptibles de transmettre l'information m.

Cette structure du dictionnaire nécessite de choisir parmi les vecteurs de $\mathcal{D}_k$ le vecteur le plus approprié à la transmission de l'information k. Nous avons établi section 9.2.3.1 que ce vecteur est le vecteur $\mathbf{d}_k^{opt}$ tel que :

- dans le cas du récepteur par filtre blanchissant,

$$opt = \arg \max_{q\in[\![0,Q-1]\!]} \tilde{\mathbf{x}}^t \tilde{\mathbf{d}}_k^q. \tag{9.91}$$

- dans le cas de l'égaliseur,

$$opt = \arg \max_{q\in[\![0,Q-1]\!]} \tilde{\mathbf{x}}^t \mathbf{d}_k^q. \tag{9.92}$$

Rappelons que ce vecteur est également celui qui sera détecté par le corrélateur si la transmission de k est correcte.

Étant donnée cette nouvelle structure du dictionnaire, les conditions de détection correcte et de robustesse établies équation (9.37) pour les besoins de la stratégie de robustesse maximale aux perturbations doivent être modifiées. Elles deviennent :

- dans le cas du récepteur par filtre blanchissant,

$$\forall m \neq k, \forall p, (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \, \tilde{\mathbf{d}}_k^{opt} > (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \, \tilde{\mathbf{d}}_m^p. \tag{9.93}$$

- dans le cas de l'égaliseur,

$$\forall m \neq k, \forall p, (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \, \mathbf{d}_k^{opt} > (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \, \mathbf{d}_m^p. \tag{9.94}$$

Transmettre k sans erreur suppose que la corrélation entre le signal reçu et $\mathbf{d}_k^{opt}$ soit supérieure à la corrélation entre le signal reçu et tous les vecteurs des sous-dictionnaires autre que $\mathcal{D}_k$ c'est à dire les $\mathbf{d}_m^q$ avec $m \neq k$. Les vecteurs du sous-dictionnaire $\mathcal{D}_k$ autre que $\mathbf{d}_k^{opt}$ ne sont pas considérés : en effet, $\mathbf{d}_k^{opt}$ est choisi pour être de corrélation maximale avec le signal audio blanchi $\tilde{\mathbf{x}}$ donc est le plus susceptible d'être de corrélation maximale avec le signal reçu (par rapport aux corrélations avec les vecteurs de $\mathcal{D}_k$ autre que $\mathbf{d}_k^{opt}$). Si ce n'est pas le cas, par exemple si un bruit de canal $\mathbf{p}$ venait à rendre la corrélation maximale avec le signal reçu pour un vecteur $\mathbf{d}_k^q$

autre que $\mathbf{d}_k^{opt}$, ce vecteur n'entraînerait pas d'erreur de détection : en effet, puisqu'il appartient à $\mathcal{D}_k$, il conduirait encore à la détection de l'information k.

La stratégie d'insertion consiste désormais à trouver le tatouage $\mathbf{t}$ donné par le signal modulé $\mathbf{v}$ et sa version filtrée $\tilde{\mathbf{v}}$ tel que :

- dans le cas du récepteur par filtre blanchissant, de trouver $\mathbf{v}$ et $\tilde{\mathbf{v}}$ tel que :

$$\begin{cases} \sigma_v^2 = \frac{\mathbf{v}^t\mathbf{v}}{N_s} \leq 1 \\ \forall m \neq k, \forall q, (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \, \tilde{\mathbf{d}}_k^{opt} \geq (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \, \tilde{\mathbf{d}}_m^q, \end{cases} \tag{9.95}$$

- dans le cas de l'égaliseur de Wiener, de trouver $\mathbf{v}$ et $\tilde{\mathbf{v}}$ tel que :

$$\begin{cases} \sigma_v^2 = \frac{\mathbf{v}^t\mathbf{v}}{N_s} \leq 1 \\ \forall m \neq k, (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \, \mathbf{d}_k^{opt} \geq (\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}})^t \, \mathbf{d}_m^q. \end{cases} \tag{9.96}$$

Les signaux $\mathbf{v}$ et $\tilde{\mathbf{v}}$ sont recherchés sous la forme d'une combinaison linéaire des vecteurs du dictionnaire $\{\mathbf{d}_m^q\}_{m\in[\![0,M-1]\!],q\in[\![0,Q-1]\!]}$. Les coefficients associés aux vecteurs du sous-dictionnaire $\mathcal{D}_k$ autre que $\mathbf{d}_k^{opt}$ ne seront pas modifiés par l'algorithme : en effet, tout comme $\mathbf{d}_k^{opt}$ ils sont porteurs de l'information k. Il n'est donc pas nécessaire de leur conférer une partie de l'énergie du signal de tatouage, puisque l'on sait que $\mathbf{d}_k^{opt}$ a le plus de chance d'être détecté.

Ce choix de coefficients de la combinaison linéaire définissant les signaux $\mathbf{v}$ et $\tilde{\mathbf{v}}$ est résolu par l'algorithme Petit Pas, en lui apportant une légère modification : le signal modulé $\mathbf{v}$ est construit itérativement en lui ajoutant de petites contributions dans la direction $\mathbf{d}_k^{opt} - \mathbf{d}_m^q$, où les indices m et q sont sélectionnés à chaque itération de sorte que :

- dans le cas du filtre blanchissant, $\tilde{\mathbf{d}}_m^q$ soit le vecteur le plus compétitif, c'est à dire le vecteur minimisant la corrélation $(\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}})^t \left(\tilde{\mathbf{d}}_k^{opt} - \tilde{\mathbf{d}}_m^q\right)$
- dans le cas de l'égaliseur de Wiener, $\mathbf{d}_m^q$ soit le plus compétitif, c'est à dire celui minimisant la corrélation $(\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}})^t \left(\mathbf{d}_k^{opt} - \mathbf{d}_m^q\right)$.

9.4.2 Performances

Nous nous proposons d'évaluer les performances du système de tatouage informé final (avec dictionnaire structuré et choix d'un tatouage de robustesse maximale aux perturbations) sur le corpus des 20 signaux. Les quatre critères établis dans le protocole expérimental sont considérés pour évaluer ces performances : l'inaudibilité, le TEB pour un canal sans perturbation, la robustesse aux perturbations non désynchronisantes et le coût en temps de calcul. Le dictionnaire sera constitué de 64 vecteurs, permettant un encodage de $K = 4$ bits par symbole et $Q = 4$ vecteurs par sous-dictionnaire. Les performances du système en boucle fermée final seront comparées aux performances du système non informé de référence établi chapitre 5, du système informé avec dictionnaire structuré seul proposé section 9.2 et du système informé basé sur l'algorithme Petit Pas développé section 9.3.

Inaudibilité

La table 9.4 présente les valeurs moyennes et des écart-types de l'ODG obtenues par tatouage des 20 signaux du corpus à un débit de 250 bit/s. Huit systèmes sont considérés correspondant à la technique d'insertion (informée ou non) et au récepteur utilisé. Ces résultats mettent en évidence une fois encore l'efficacité du système de tatouage informé final en terme d'inaudibilité. Le tatouage obtenu, malgré l'introduction de la copie locale du récepteur à l'émetteur et la

Système	non informé de référence	informé		
		dictionnaire structuré	Petit Pas	final
Filtre blanchissant	$-1,0$ $(0,07)$	$-1,02$ $(0,06)$	$-1,01$ $(0,06)$	$-1,03$ $(0,06)$
Égaliseur de Larbi	$-1,0$ $(0,07)$	$-1,18$ $(0,10)$	$-1,01$ $(0,06)$	$-1,01$ $(0,08)$

Table 9.4 – Valeurs moyennes de l'ODG et écart-type (entre parenthèses) sur le corpus de 10 signaux audio pour deux récepteurs et quatre systèmes de tatouage informés et non.

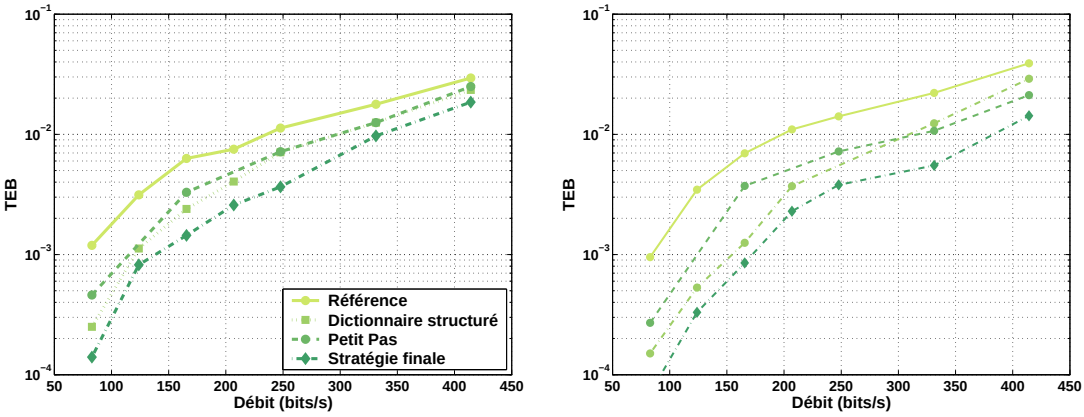


Figure 9.14 – Comparaison des TEBs du système informé final avec les systèmes non informés et informés présentés précédemment pour deux récepteurs : (a) le filtre blanchissant et (b) l'égaliseur de Larbi. Le canal ne subit pas de perturbations.

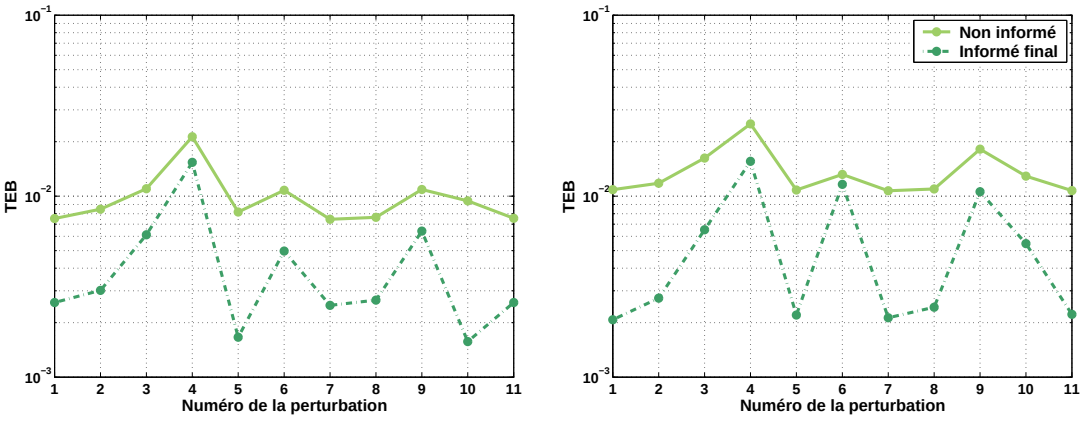
modification de l'étape de modulation, est audible mais non gênant, comme attendu par le paramétrage du module de contrôle d'inaudibilité.

TEBs pour un canal sans perturbation

La figure 9.14 présente les TEBs du système de tatouage informé final en fonction du débit pour les deux types de récepteurs (et donc de copie locale du récepteur à l'émetteur) dans le cas où le canal ne subit pas de perturbations. Cette figure rappelle à titre de comparaison les performances du système non informé et des deux systèmes informés établis précédemment, à savoir celui avec dictionnaire structuré seul et celui avec l'algorithme Petit Pas seul.

Ces courbes mettent en évidence l'intérêt de combiner les deux stratégies d'insertion informée : en effet, les TEBs du système informé final sont inférieurs aux TEBs de chacune des deux stratégies. Ce système final offre donc une très nette amélioration des performances par rapport au système non informé : en effet, dans le cas d'une réception par filtre blanchissant, les TEBs sont quasiment divisés par 10 à bas débit (50 bit/s) et par $1,2$ à débit élevé (à 400 bit/s). Dans le cas de l'égaliseur de Larbi, cette amélioration est encore plus nette : les TEBs sont divisés par 10 jusqu'à environ 150 bit/s et par 3 à 400 bit/s.

Grâce à l'apport des stratégies d'insertion informée, une transmission de TEB 10^{-3} et d'ODG -1 peut désormais être obtenue à 140 bit/s si le récepteur est basé sur le filtre blanchissant et à



(a) (b)

Figure 9.15 – TEBs du système avec stratégie d'insertion informée final et du système non informé en fonction des différentes perturbations de canal (représentées par leur numéro) lorsque le débit de transmission est fixé à 206 bit/s. Le récepteur est : (a) le filtre blanchissant et (b) l'égaliseur de Larbi

170 bit/s si le récepteur est l'égaliseur de Larbi. Avec le système non informé, cette transmission n'était envisageable qu'à 80 bit/s soit une augmentation par 2 du débit de transmission possible.

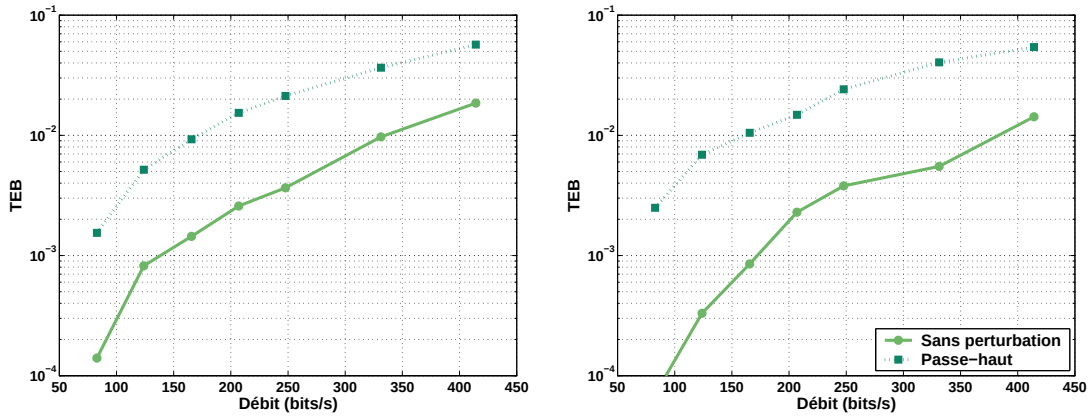
Robustesse aux perturbations

Les figures 9.15 présentent l'évolution du TEB du système informé final à 206 bit/s face aux différentes perturbations non désynchronisantes considérées. Ces perturbations sont rappelées table 9.5. Les deux récepteurs du système de tatouage sont considérés : la figure (a) est dédiée au récepteur par filtre blanchissant et la figure (b) à l'égaliseur de Wiener. Les TEBs du système non informé sont également rappelés à titre de comparaison.

Numéro	Perturbations	Numéro	Perturbations
1	Sans perturbations	**7**	Sous échantillonnage (22050 kHz)
2	Compression MPEG 96 kbit/s	**8**	Quantification (12 bits)
3	Compression MPEG 64 kbit/s	**9**	Ajout d'un bruit blanc (30 dB)
4	Filtrage passe-haut* (250 Hz)	**10**	Amplification (réduction de 50%)
5	Filtrage passe-bas* (9 kHz)	**11**	Compression de dynamique*
6	Ajout d'écho		

Table 9.5 – Liste des perturbations et de leur numéro associé. Les perturbations indiquées par une * sont réalisées par le Stirmark Audio.

L'amélioration des TEBs liés à la stratégie d'insertion informée se répercute naturellement sur les TEBs obtenus pour les différentes perturbations. Ces TEBs sont meilleurs dans le cas informé que dans le cas non informé. Les opérations MPEG à débit élevé (2), le filtrage passe-bas (5), les modifications de dynamiques (10 et 11) modifient peu les TEBs ; les compressions MPEG (3), ajout d'écho (6) et ajout de bruit (9) ont comme dans le cas du système non informé les perturbations les plus dégradantes. En particulier, l'ajout de bruit multiplie par $2,3$ les TEBs du



(a) (b)

Figure 9.16 – TEBs en fonction du débit obtenus par le système informé lorsque le canal ne subit pas de dégradation et lorsqu'il est soumis à un filtre passe-haut. Le récepteur utilisé est : (a) le filtre blanchissant et (b) l'égaliseur de Larbi

	Émetteur	**Récepteur**
Non informé	78%	117%
Informé avec Filtre Blanchissant	2030%	590 %
Informé avec Égaliseur de Larbi	4000%	47 %

Table 9.6 – Coût en temps de calcul par rapport au temps-réel pour les différentes configurations d'émission et de réception.

système avec filtre blanchissant et par 5 ceux du système avec égaliseur de Wiener par rapport au cas sans perturbation. La perturbation par filtrage passe-haut est, comme dans le cas du système informé, la perturbation la plus dégradante : elle introduit une augmentation des TEBs par un facteur 6 pour le filtre blanchissant et un facteur $7,5$ pour l'égaliseur de Wiener, soit près de deux fois plus que la majorité des perturbations ! Cette perturbation devra donc faire l'objet d'une étude à part entière pour essayer d'en limiter les effets sur les performances du système. Pour autant, les TEBs du système informé restent inférieurs aux TEBs du système non informé, quelle que soit la perturbation considérée.

Pour conclure cette étude sur la robustesse du système informé aux perturbations, les TEBs obtenus dans le cas de la perturbation la plus dégradante pour le système (*i.e* le filtrage passe-haut) ont donc été tracé en fonction du débit pour les deux récepteurs considérés. La figure 9.16 présente les résultats obtenus et les comparent aux TEBs du système sans perturbation. Compte tenu de la dégradation introduite, une transmission de TEB 10^{-3} peut être atteinte à environ 75 bit/s quelle que soit la perturbation non désynchronisante considérée (contre 50 bit/s dans le cas non informé).

Coût en temps de calcul

La table 9.6 présente à titre indicatif les coûts en temps de calcul des différentes configurations de système. Ce coût est le ratio entre le temps nécessaire au calcul du tatouage et la durée de

signal traité sur un PC dont les caractéristiques sont données dans le protocole expérimental. Partant d'un système non informé quasi temps-réel, l'introduction à l'émetteur des copies locales des récepteurs pour les besoins de la stratégie d'insertion informée augmente très nettement le coût en temps de calcul de la chaîne : cette augmentation est liée à l'estimation des signaux intervenant dans le procédé de détection. Elle est d'autant plus importante dans le cas de la réception par égaliseur de Wiener, où les opérations de filtrage sont nettement plus nombreuses que dans le cas du récepteur par filtre blanchissant. En effet, les filtres doivent être spécifiquement calculés pour chaque signal de tatouage construit par l'algorithme. Le coût du récepteur pour les systèmes informés augmente également par rapport à celui du système non informé. Cette augmentation est prévisible puisque l'introduction d'un dictionnaire structuré conduit à augmenter le nombre M de vecteurs du dictionnaire par rapport au cas non informé. Dans le cas du filtre blanchissant, elle se traduit par une augmentation relativement importante, puisque nous avons établi section 7.2.4 que le coût de ce récepteur croît exponentiellement avec M. Dans le cas de l'égaliseur de Larbi, dont l'implémentation est peu sensible à la taille du dictionnaire, ce coût augmente peu et permet d'envisager une application temps-réel du récepteur.

Conclusion

Dans ce chapitre, nous nous sommes intéressés à la construction de stratégies d'insertion informée adaptées à la structure en boucle fermée du système de tatouage et à un schéma de modulation M-aire. Ces stratégies ont pour objectif de combler deux problèmes inhérents à la construction du système de tatouage non informé et à l'origine d'erreurs de transmission. La première propose de structurer le dictionnaire d'émission suivant le modèle de Costa : le nombre de vecteurs utilisés pour la transmission d'un même symbole k est accru de sorte à augmenter les chances de trouver une région de détection de k à proximité du signal audio. La seconde stratégie propose pour un dictionnaire non structuré de choisir le signal de tatouage de robustesse maximale aux perturbations sous contrainte d'inaudibilité. Cette robustesse se traduit soit sous la forme d'un paramètre de robustesse à maximiser, soit sous la forme d'une expression de la probabilité d'erreur face à la perturbation, donnant lieu à deux problèmes d'optimisation sous contraintes. Trois algorithmes ont été proposés pour résoudre ce problème. Un seul, basé sur un algorithme itératif dit Petit Pas, a été retenu en tant que meilleur compromis entre performances et coût en temps de calcul. Les performances de ces deux stratégies ont été évaluées sur un modèle théorique de la chaîne de tatouage et sur le système de tatouage en boucle fermée pour les deux schéma de réception. Ces performances montrent une nette amélioration des TEBs du système, permettant d'envisager une transmission à 10^{-3} pour un débit de l'ordre de 125 bit/s (contre 80 bit/s dans le cas d'un système non informé). Pour finir, ces deux stratégies ont été combinées sous une unique solution, appelée stratégie d'insertion avec dictionnaire structuré et choix d'un tatouage de robustesse maximale aux perturbations. Cette stratégie bénéficie des apports de chaque stratégie sur les performances du système et permet d'améliorer plus encore les TEBs du système. Ainsi une transmission de TEB 10^{-3} peut être obtenue à 170 bit/s (avec pour récepteur l'égaliseur de Larbi) lorsque le canal ne subit pas de perturbations. Cette transmission est relativement robuste aux perturbations de canal avec une augmentation inférieure à 3% des TEBs pour la majorité des perturbations. Une perturbation par filtrage passe-haut s'est révélée être la perturbation la plus dégradante puisqu'elle provoque une augmentation des TEBs de $6,5\%$; une étude à part entière devra être entreprise pour limiter ses effets sur les performances du système. Pour autant, une transmission imposée à 10^{-3} peut être maintenue quelle que soit la perturbation non désynchronisante à condition de choisir un débit d'environ 75 bit/s (contre

50 bit/s pour le système non informé).

Cette stratégie d'insertion informée, efficace, est adapté au modulation M-aire. Nous allons donc nous intéressé dans le chapitre qui suit au cas des modulations CDMA.

Chapitre 10

Tatouage CDMA informé

Dans ce chapitre, nous nous intéressons plus particulièrement au système de tatouage lorsque la modulation est une modulation CDMA. Nous proposons d'établir une stratégie d'insertion informée basée comme précédemment sur l'insertion d'une copie locale du récepteur à l'émetteur. Cette copie locale du récepteur permet de rétroagir sur le choix du signal modulé et plus particulièrement des facteurs d'amplitude caractérisant l'énergie consacrée à la transmission de chaque bit. Ces amplitudes sont choisies de sorte à maximiser la robustesse de la transmission à un bruit de perturbation.

10.1 De la nécessité d'une stratégie d'insertion CDMA informée

Considérons la transmission du K-uplet binaire $\mathbf{b}$ pendant le l-ième temps symbole. Le signal modulé, de part le schéma de modulation CDMA, s'exprime sous la forme d'une combinaison linéaire des vecteurs du dictionnaire d'émission :

$$\mathbf{v} = \sum_{m=0}^{K-1} \gamma_m (2\mathbf{b}[m] - 1)\mathbf{d}_m = \mathcal{D}\left(\gamma \otimes (2\mathbf{b} - 1)\right), \tag{10.1}$$

où $\gamma = [\gamma_0, .., \gamma_{K-1}]$ désigne les facteurs d'amplitude caractérisant l'énergie dédiée à la transmission de chaque bit du K-uplet $\mathbf{b}$ et $\otimes$ le produit terme à terme. Pour une stratégie non informée, chaque bit est transmis avec la même énergie, de sorte que γ_m est choisi égal à $\frac{1}{\sqrt{K}}$ pour satisfaire à la contrainte de puissance imposée par le filtre de mise en forme perceptuelle $H(f)$ au signal modulé. La stratégie d'insertion informée remet en cause le choix d'une énergie γ_m indépendante du signal audio.

Pour exploiter la connaissance *a priori* du signal audio durant l'émission du tatouage, nous avons choisi d'introduire une copie locale du récepteur à l'émetteur dont un modèle unifié a été proposé section 8.3. Ce modèle formalise les conditions d'une transmission CDMA inaudible et correcte, qui déterminent le choix des énergies γ_m adaptées maximisant la robustesse de la transmission aux perturbations.

10.1.1 Formalisation des conditions de transmission

Les conditions d'inaudibilité, établies section 8.2.1, sont liées à la contrainte de puissance imposée au signal modulé par le mise en forme perceptuelle $H(f)$ et le gain α. Étant donnée l'expression de $\mathbf{v}$, elles s'expriment sous la forme :

$$\sigma_v^2 = \frac{1}{N_s}\mathbf{v}^t\mathbf{v} = \gamma^t\gamma \leq 1. \tag{10.2}$$

Les conditions d'une détection correcte et robuste du K-uplet binaire $\mathbf{b}$ transmis dépendent du détecteur par décorrélation que nous avons choisi. La détection est basée sur le calcul de la corrélation $\tilde{\mathbf{c}}$ entre le signal reçu $\mathbf{x}+\mathbf{t}+\mathbf{p}$ et les vecteurs du dictionnaire de réception. Cette corrélation est ensuite pondérée par la matrice d'autocovariance du dictionnaire de réception, pour annuler les interférences entre les vecteurs du dictionnaire. La détection tient finalement aux signes des composantes du vecteur de corrélation $\tilde{\mathbf{c}}$ ainsi obtenu. Nous avons déjà établi section 8.2.2.2 l'expression de ces conditions de réception : pour garantir une transmission correcte de $\mathbf{b}$, le signe de chaque composante de $\tilde{\mathbf{c}}'$ doit être identique au signe des composantes de $(2\mathbf{b}-1)$:

$$(2\mathbf{b}-1)\otimes\tilde{\mathbf{c}}' > 0. \tag{10.3}$$

Appliquons maintenant ces résultats au modèle de la chaîne de tatouage donné figure 8.3. En substituant à $\mathbf{t}$ son expression en fonction du K-uplet binaire émis, à savoir $\mathbf{t} = \alpha\mathbf{v} = \alpha\mathcal{D}\left(\gamma\otimes(2\mathbf{b}-1)\right)$, ce vecteur de corrélation peut être écrit sous la forme :

$$\tilde{\mathbf{c}} = \mathcal{D}^t\left(\mathbf{x}+\mathbf{t}+\mathbf{p}\right) = \mathcal{D}^t\mathbf{x} + \alpha R_{\mathcal{D}}\left(\gamma\otimes(2\mathbf{b}-1)\right) + \mathcal{D}^t\mathbf{p}, \tag{10.4}$$

faisant apparaître la matrice d'autocorrélation empirique du dictionnaire $\mathcal{R}_{\mathcal{D}}$. Ce vecteur est ensuite pondéré par $\mathcal{R}_{\mathcal{D}}^{-1}$; il en est déduit que :

$$\tilde{\mathbf{c}}' = \mathcal{D}^\dagger\mathbf{x} + \alpha\left(\gamma\otimes(2\mathbf{b}-1)\right) + \mathcal{D}^\dagger\mathbf{p}, \tag{10.5}$$

où $\mathcal{D}^\dagger = R_{\mathcal{D}}^{-1}\mathcal{D}$ désigne la pseudo-inverse de $\mathcal{D}$. Le K-uplet reçu est ensuite décidé en fonction du signe de $\tilde{\mathbf{c}}'$. Nous déduisons de (10.3) et (10.5) les $M = K$ inégalités traduisant le choix adapté de γ pour la transmission correcte et robuste du K-uplet binaire sous forme vectorielle :

$$(2\mathbf{b}-1)\otimes\mathcal{D}^\dagger\mathbf{x} + \alpha\gamma + (2\mathbf{b}-1)\otimes\mathcal{D}^\dagger\mathbf{p} > 0. \tag{10.6}$$

La stratégie d'insertion informée consiste alors à déterminer les coefficients γ satisfaisant à la contrainte d'inaudibilité (10.2) et aux conditions de détection (10.6) pour un bruit de perturbation $\mathbf{p}$ maximal.

10.1.2 Interprétation géométrique

L'expression des conditions de transmission correcte et robuste donnée équation (10.6) montre la très forte dépendance de la détection au signal audio. Pour mettre en évidence cette dépendance, reprenons l'interprétation géométrique proposée par Miller [MCB00] et sa représentation dans le cas d'un dictionnaire de $M = 2$ vecteurs orthogonaux $\mathcal{D} = \{\mathbf{d}_0, \mathbf{d}_1\}$, en supposant pour l'occasion le bruit de canal $\mathbf{p}$ nul.

Supposons que le 2-uplet à émettre soit $\mathbf{b} = [1; 1]$. Le signal modulé choisi par le système de tatouage non informé est donc $\mathbf{v} = \frac{1}{\sqrt{2}}(\mathbf{d}_0 + \mathbf{d}_1)$ tandis que le signal choisi par la stratégie d'insertion informée prend la forme : $\mathbf{v} = \gamma_0\mathbf{d}_0 + \gamma_1\mathbf{d}_1$.

Pour une modulation CDMA, l'espace de détection est l'espace engendré par les vecteurs de $\mathcal{D}^\dagger$. Dans le cas particulier du modèle de la chaîne de tatouage, le dictionnaire d'émission est identique au dictionnaire de réception, de sorte que $\mathcal{D}^\dagger = \mathcal{D}$. L'espace de détection dans lequel nous nous placerons est donc l'espace engendré par les vecteurs du dictionnaire.

Dans cet espace, la région d'inaudibilité, notée $\mathfrak{R}_i$, définie comme étant l'ensemble des signaux audio tatoué $\mathbf{y} = \mathbf{x} + \mathbf{t}$ avec $\mathbf{t} = \alpha\mathbf{v}$ satisfaisant la contrainte d'inaudibilité (10.2) peut être représentée par le disque de rayon α centré sur $\mathbf{x}$.

Étant donné le démodulateur par décorrélation choisi, la détection de chaque bit du 2-uplet binaire peut être effectuée indépendamment des autres bits. Le bit porté par $\mathbf{d}_0$ est correctement détecté si sa corrélation avec le signal reçu est du même signe que $\mathbf{b}[0]$. En particulier la région de détection du bit '1' porté par $\mathbf{d}_0$, notée $\mathfrak{R}_{\mathbf{d}_0}$, est le demi-plan supérieur dirigé par $\mathbf{d}_0$ et présenté figure 10.1 par des hachures. De la même façon, le bit porté par $\mathbf{d}_1$ est correctement détecté si la corrélation avec le signal reçu est du même signe que $\mathbf{b}[1]$. La région de détection du bit '1' porté par $\mathbf{d}_1$, notée $\mathfrak{R}_{\mathbf{d}_1}$, est le demi-plan supérieur dirigé par $\mathbf{d}_0$ et également présenté figure 10.1 par des hachures. Les deux bits du 2-uplet sont correctement détectés si le signal reçu est à l'intersection des deux régions de détection établi précédemment. Cette région, notée $\mathfrak{R}_d$ est donc le quart de plan dirigé par $(\mathbf{d}_0 + \mathbf{d}_1)$.

Le tatouage doit donc être choisi à l'intersection de la région d'inaudibilité et de la région de détection correcte $\mathfrak{R}_d$.

Quatre configurations de signal audio, présentées figure 10.1, sont considérées. Dans la première, donnée figure 10.1 (a), la projection $\mathbf{x}$ du signal audio dans l'espace de détection est à elle seule porteuse de l'information $\mathbf{b}$: en effet, elle se situe dans la région de détection correcte $\mathfrak{R}_d$ des 2 bits de $\mathbf{b}$. Il n'est donc pas nécessaire d'ajouter un tatouage $\mathbf{t}$ pour garantir la transmission correcte de l'information.

La seconde et la troisième configuration, présentée respectivement figure 10.1 (b) et (c), présentent un cas moins favorable : la projection du signal audio ne suffit plus à transmettre correctement les 2 bits d'information. Si aucun tatouage n'est ajouté, le bit porté par $\mathbf{d}_1$ est correctement détecté puisque $\mathbf{x}$ est dans le région de détection $\mathfrak{R}_{\mathbf{d}_1}$; par contre le bit porté par $\mathbf{d}_0$ est erroné. L'ajout d'un tatouage devient donc nécessaire. Dans le cas de la seconde configuration, le tatouage choisi par le système non informé, i.e. $\mathbf{t}_{ni} = \frac{\alpha}{\sqrt{2}}(\mathbf{d}_0 + \mathbf{d}_1)$, permet d'amener le signal reçu $\mathbf{x} + \mathbf{t}$ dans la région de détection des deux bits $\mathfrak{R}_d$. Dans le cas de la troisième configuration, ce tatouage ne suffit pas à compenser la forte corrélation du signal audio dans la direction $-\mathbf{d}_0$. Pourtant, l'intersection entre la région d'inaudibilité et la région de détection correcte n'est pas vide. Il existe donc un tatouage autre que le tatouage $\mathbf{t}_{ni}$ qui permette la réception correcte des deux bits d'information. Ce tatouage est en fait $\mathbf{t}_i = \alpha\mathbf{d}_0$: le signal audio étant à lui seul porteur du bit '1' dans la direction $\mathbf{d}_1$, il n'est donc pas nécessaire de dédier une partie de la puissance du tatouage à transmettre ce bit. Par contre, cette puissance peut être consacrée à transmettre le bit '1' dans la direction $\mathbf{d}_0$ et combattre plus efficacement le bruit audio que ne le ferait le tatouage non informé.

La dernière configuration en figure 10.1 (d) présente le cas de plus défavorable, dans lequel aucun tatouage $\mathbf{t}$ limité par la contrainte d'inaudibilité ne permet de transmettre le bit '1' dans la direction $\mathbf{d}_0$. Dans ce cas de figure, une stratégie d'abandon peut être envisagée : sachant que

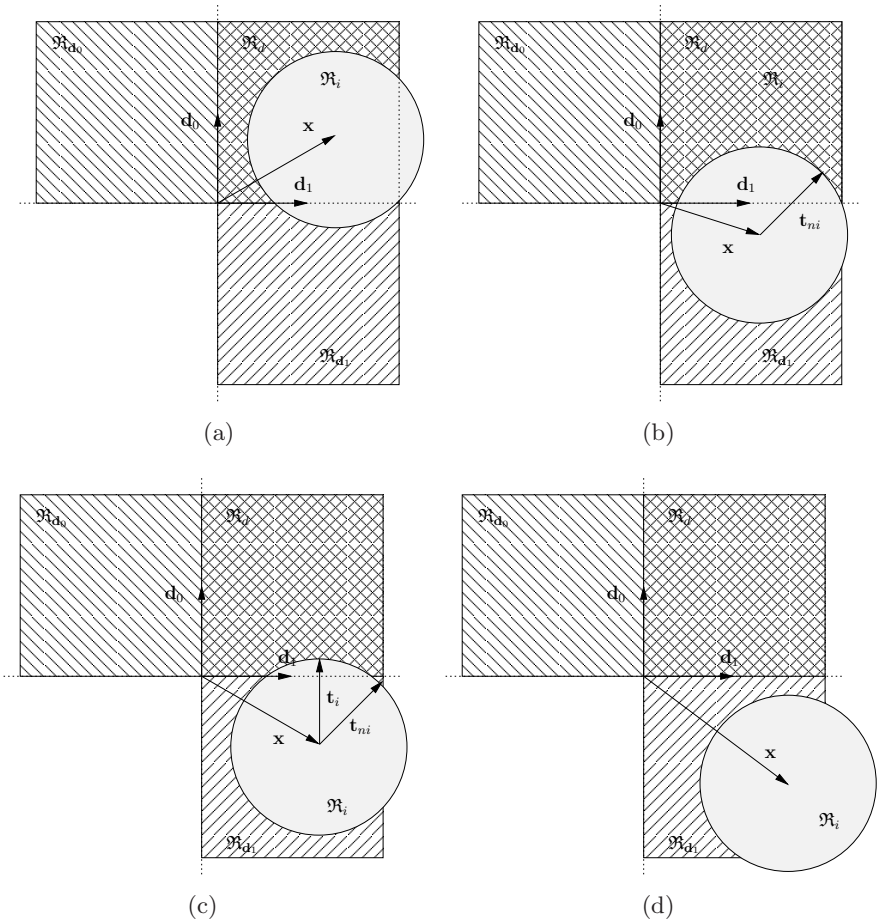


Figure 10.1 – Mise en évidence des régions de distorsion acceptable et de détection pour différentes configurations de signaux de la plus favorable à la moins favorable

le bit reçu dans la direction $\mathbf{d}_0$ sera de toute façon erroné, toute la puissance du tatouage peut être consacrée à maximiser la robustesse de la transmission du bit dans la direction $\mathbf{d}_1$.

Ces visualisations mettent en évidence la nécessité de choisir les coefficients γ de façon adaptée au signal audio ; ces coefficients seront adaptés si le tatouage qui en résulte est inaudible et garantit la transmission correcte et robuste de l'information $\mathbf{b}$ transmise.

10.2 Stratégie d'insertion CDMA informée

10.2.1 Vers la prise en compte du bruit de canal

La stratégie d'insertion informée consiste à trouver le signal modulé $\mathbf{v} = \mathcal{D}(\gamma \otimes (2\mathbf{b} - 1))$ porteur de l'information $\mathbf{b}$ qui satisfait la contrainte d'inaudibilité donnée par l'équation (10.2) et la transmission correcte et robuste de l'information, sous la forme des M inégalités de l'équation (10.6). Ces dernières font intervenir un bruit de canal $\mathbf{p}$ que l'on souhaite maximal mais qui n'est pas connu lors de la phase d'insertion. Nous proposons donc d'introduire un

paramètre de robustesse au bruit σ_p^2 de manière similaire à la stratégie proposée dans [LDC04] et à la stratégie M-aire basée sur l'approche déterministe [1] présentée section 9.3.2. Ce paramètre de robustesse que l'on souhaite maximiser se substitue au vecteur de bruit $(2\mathbf{b}-1)\otimes\mathcal{D}^\dagger\mathbf{p}$ dans l'équation (10.6). Les M inégalités traduisant la détection robuste du M-uplet binaire en sont déduites :

$$\alpha\gamma + (2\mathbf{b}-1)\otimes\mathcal{D}^\dagger\mathbf{x} \geq \tau_p^2[1\cdots 1]^t, \tag{10.7}$$

avec τ_p^2 maximum, où $[1\cdots 1]^t$ désigne le vecteur colonne $M\times 1$ composé de 1.

L'objectif de la stratégie d'insertion informée est alors de déterminer pour le M-uplet binaire à transmettre $\mathbf{b}$ le vecteur d'amplitude γ qui satisfait la contrainte d'inaudibilité (10.2) et les conditions d'une détection robuste (10.7) avec un paramètre de robustesse au bruit τ_p^2 maximal. Si une telle solution γ est trouvée, l'équation (10.7) montre que la valeur maximum de τ_p^2 est donnée par la plus petite composante du vecteur $\alpha\gamma + (2\mathbf{b}-1)\otimes\mathcal{D}^\dagger\mathbf{x}$:

$$\tau_p^2 = \min\Big(\alpha\gamma + (2\mathbf{b}-1)\otimes\mathcal{D}^\dagger\mathbf{x}\Big). \tag{10.8}$$

10.2.2 Choix du vecteur d'amplitude γ

Le choix du vecteur d'amplitude optimal γ, satisfaisant à la fois (10.7) et (10.7) avec un paramètre de robustesse σ_q^2 donné par (10.8) maximal, est lié à un problème d'optimisation sous contrainte pouvant être formalisé par :

$$\begin{cases} \gamma = \arg\max_\mu \tau_p^2(\mu) \\ \tau_p^2(\mu) = \min\big(\alpha\mu + (2\mathbf{b}-1)\otimes\mathcal{D}^\dagger\tilde{\mathbf{x}}\big) \\ \mu^t\mu = 1 \end{cases} \tag{10.9}$$

Ce problème d'optimisation peut être résolu par un algorithme itératif, proche de l'algorithme Petit Pas présenté section 9.3.2.3. Il consiste à incrémenter pas à pas les composantes de γ en sélectionnant celle dont la transmission est la plus défavorable (et donc requière le plus d'énergie). Les étapes de cet algorithme sont décrites dans l'algorithme 10.1.

10.2.3 Performances théoriques

Plusieurs critères sont envisagés pour évaluer les performances de la stratégie d'insertion informée. Le premier est bien évidemment le TEB. La figure 10.2 présente les TEBs obtenus en fonction du paramètre N_bRSB dans deux configurations. La première, donnée figure (a), rappelle les TEBs de la chaîne CDMA non informée tracés à partir de l'expression théorique de la probabilité d'erreur établie section 4.4.2. Ces TEBs sont, rappelons-le, indépendants du nombre $M = K$ de vecteurs du dictionnaire. La seconde configuration, donnée figure (b), présente les TEBs de la stratégie CDMA informée pour différentes valeurs de $M = K$. Ces TEBs sont obtenus par simulation du modèle de la chaîne de tatouage avec 10^6 bits. On constate que la stratégie d'insertion informée permet effectivement d'améliorer les TEBs : contrairement au cas

1. De la même façon que pour les stratégies d'insertion M-aires informées, nous aurions pu modéliser le bruit de canal $\mathbf{p}$ comme la réalisation d'un vecteur aléatoire dont les composantes suivent une loi $\mathcal{N}(0,\sigma_p^2)$. Cette modélisation nous aurait permis d'évaluer la probabilité d'erreur de transmission et de déduire une stratégie proche de l'approche probabiliste exposée pour les modulations M-aires. Néanmoins, les résultats des modulations M-aire laisse à penser que cette approche aurait conduit à la même robustesse du système aux perturbations que l'algorithme proposé ici mais à un coût en temps de calcul nettement plus élevé.

Étape 1 : $\gamma^{(0)}$ est initialisé à zéro de sorte que le signal modulé à l'itération 0, $\mathbf{v}^{(0)}$, est initialisé au vecteur nul.

Étape 2 : Notons $\gamma^{(n)}$ le vecteur des coefficients à la n-ième itération et $\mathbf{v}^{(n)} = \mathcal{D}\left(\gamma^{(n)} \otimes (2\mathbf{b}-1)\right)$ le signal modulé qui lui est associé. Étant donné $\gamma^{(n)}$, la plus petite composante de $\gamma + (2\mathbf{b}-1) \otimes \tilde{\mathcal{D}}^{\dagger}\tilde{\mathbf{x}}$ est sélectionnée. Notons $comp$ son indice. Cette composante indique que la transmission du bit b_{comp} dans la direction $\mathbf{d}_{comp}$ est la moins robuste et donc la plus susceptible d'induire une erreur de détection. Le facteur d'amplitude γ_{comp} caractérisant l'énergie consacrée à la transmission du bit b_{comp} est donc augmenté d'un pas ρ.

$$\begin{cases} \forall m \neq comp, \gamma_m^{(n+1)} = \gamma_m^{(n)} \\ \gamma_{comp}^{(n+1)} = \gamma_{comp}^{(n)} + \rho \end{cases} \tag{10.10}$$

Si la puissance de $\mathbf{v}^{(n+1)} = \mathcal{D}\left(\gamma^{(n+1)} \otimes (2\mathbf{b}-1)\right)$ est inférieure à 1, l'étape 2 est réitérée. Sinon, le signal modulé optimal est obtenu en normalisant $\mathbf{v}^{(n+1)}$ en puissance :

$$\mathbf{v} = \frac{\mathbf{v}^{(n+1)}}{\sigma_{\mathbf{v}^{(n+1)}}} \text{ avec } \sigma^2_{\mathbf{v}^{(n+1)}} = \frac{\mathbf{v}^{(n+1)\,t}\,\mathbf{v}^{(n+1)}}{N_s}. \tag{10.11}$$

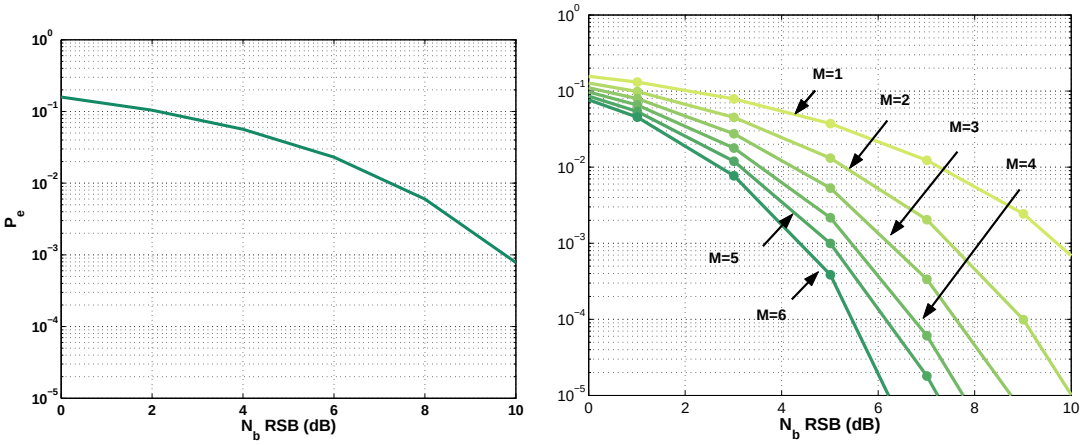
Algorithme 10.1: Algorithme de choix des amplitudes pour les modulations CDMA informées.

du système non informé, le TEB du système informé est, cette fois, d'autant meilleur que le nombre de vecteurs du dictionnaire K est grand.

Pour apprécier l'amélioration des performances, trois autres critères sont proposés figure 10.3 lorsque le dictionnaire est constitué de $K = 4$ vecteurs. Le premier, présenté figure 10.3 (a), est un Critère d'Amplitude (CA) qui permet d'apprécier le facteur d'amplitude moyen consacré à la transmission de chaque bit. Ce critère est la moyenne des facteurs d'amplitude $\{\gamma_k\}_{k\in[\![0,K-1]\!]}$:

$$CA = \frac{1}{K}\sum_{k=0}^{K-1}\gamma_k. \tag{10.12}$$

Dans le cas du système non informé, l'énergie consacrée à la transmission de chaque bit est constante pour ne pas favoriser une transmission par rapport à une autre : le critère d'amplitude est donc constant quel que soit le paramètre N_bRSB et égale à $CA = \frac{1}{\sqrt{K}}$. Dans le cas du système informé, les amplitudes γ_k sont adaptées au signal audio. La figure (a) montre que plus ce signal est puissant, plus la valeur moyenne de γ_k est faible devant $\frac{1}{\sqrt{K}}$. En effet, dans le cas de configurations de signaux audio défavorables, la transmission de certains bits du K-uplet est privilégiée par rapport aux autres. Rappelons notamment le cas de la configuration présentée figure 10.1 (c) ; la transmission du bit b_0 pour être correcte nécessitait de lui associer un facteur d'amplitude γ_k très important au détriment de la transmission du bit b_1, pour laquelle l'amplitude choisie était proche de 0. Si une même configuration de transmission se produit maintenant dans le cas de $K = 4$ vecteurs, le critère d'amplitude est dans le pire des cas $CA = \frac{1}{K} = \frac{1}{4}$.



(a) (b)

Figure 10.2 – TEBs en fonction du paramètre N_bRSB dans le cas du système non informé (a) et dans le cas de la stratégie d'insertion informée avec modulation CDMA (b) pour une taille M du dictionnaire variable.

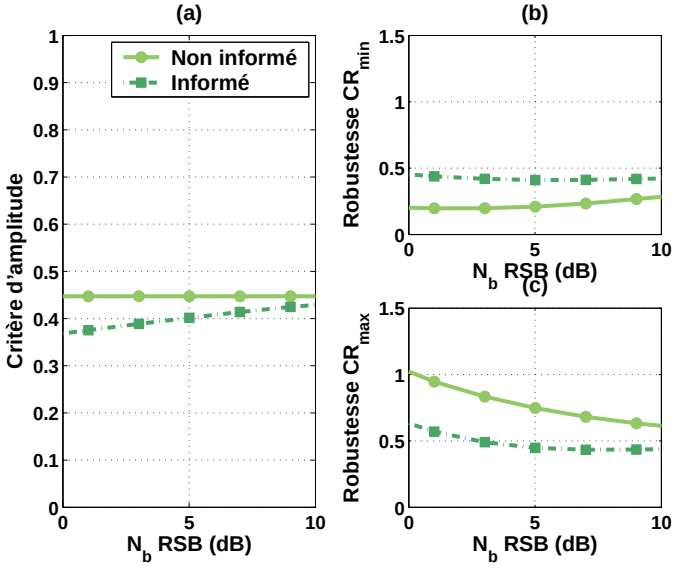


Figure 10.3 – Comparaison de la stratégie d'insertion informée et de la stratégie non informée pour les modulations CDMA à travers 3 critères : (a) le critère de puissance, (b) le critère de robustesse minimale et (c) le critère de robustesse maximale en fonction du paramètre N_bRSB.

Lorsque la puissance du signal audio décroît, on constate que le critère d'amplitude du système informé tend vers celui du système non informé. En effet, le signal audio ne dégradant que peu la transmission, les conditions de transmission de chaque bit du K-uplet binaire sont peu différentes et nécessitent une même amplitude proche de $\frac{1}{\sqrt{K}}$.

Le deuxième et le troisième critère s'attachent à évaluer la robustesse du système aux bruits de

canal. Cette robustesse dépend des M inégalités données équation (10.7) traduisant la détection correcte de l'information $\mathbf{b}$ pour un paramètre de robustesse τ_p^2 modélisant la projection du bruit de canal $\mathbf{p}$ sur le dictionnaire de réception. La robustesse minimum de la transmission peut donc être évaluée par la mesure de τ_p^2, donnant lieu au critère CR_{min} :

$$CR_{min} = \tau_p^2 = \min\left(\alpha\gamma + (2\mathbf{b} - 1) \otimes \mathcal{D}^\dagger \tilde{\mathbf{x}}\right). \tag{10.13}$$

Tant que le bruit de canal $\mathbf{p}$ correspond à un paramètre de robustesse inférieur à τ_p^2, la réception de l'ensemble des bits du K-uplet binaire peut être supposée correcte. Lorsque le bruit dépasse τ_p^2, certains des bits du K-uplet seront erronés ; ceux dont la valeur de $\gamma + (2\mathbf{b} - 1) \otimes \tilde{\mathcal{D}}^\dagger \tilde{\mathbf{x}}$ est plus grande que τ_p^2 pourront être supposés sans erreur, et ce tant que le paramètre de robustesse associé au bruit ne dépasse pas

$$CR_{max} = \max\left(\alpha\gamma + (2\mathbf{b} - 1) \otimes \mathcal{D}^\dagger \tilde{\mathbf{x}}\right), \tag{10.14}$$

où max désigne ici la plus grande composante du vecteur $\alpha\gamma + (2\mathbf{b} - 1) \otimes \mathcal{D}^\dagger \tilde{\mathbf{x}}$. Ce critère indique la robustesse maximum au delà de laquelle la transmission de tous les bits du K-uplet peuvent être supposé erronés. Les figures 10.3 (b) et (c) représentent les valeurs moyennes du critère de robustesse minimum et maximum en fonction du paramètre N_bRSB pour la stratégie d'insertion informée et celle non informée. Elles mettent en évidence que la stratégie d'insertion informée permet d'accroître la robustesse minimum du système au détriment de la robustesse maximum. En effet, l'augmentation de l'amplitude des bits dont la transmission est la plus défavorable conduit à diminuer l'amplitude des bits dont la transmission est la plus favorable par rapport à l'amplitude $\frac{1}{\sqrt{K}}$ qui leur est attribuée par le système non informé : de fait, la robustesse de la transmission de ces bits est diminué. Le système d'insertion informée risque donc d'être moins robuste aux bruits de canal que le système non informé lorsque ce bruit est important.

10.3 Application au système de tatouage en boucle fermée

10.3.1 Principe

Comme dans le cas des modulations M-aire, appliquer la stratégie d'insertion informée développée pour le modèle de la chaîne au système de tatouage en boucle fermée est aisée : la copie locale du récepteur permet d'estimer les signaux intervenant dans le procédé de détection. Il s'agit du signal audio filtré $\tilde{\mathbf{x}}$, du signal modulé filtré $\tilde{\mathbf{v}}$ et d'un dictionnaire $\tilde{\mathcal{D}} = \{\tilde{\mathbf{d}}_m\}$ contenant l'ensemble des vecteurs du dictionnaire d'émission $\mathcal{D}$ filtrés par les filtres de la chaîne. Ce dictionnaire est une estimation du dictionnaire de réception lorsque le récepteur est le filtre blanchissant. Au regard de ces signaux, la stratégie d'insertion informée peut être formalisée par la recherche du signal modulé $\mathbf{v}$ et de sa version filtrée $\tilde{\mathbf{v}}$ satisfaisant aux équations suivantes :

- dans le cas du filtre blanchissant,

$$\begin{cases} \sigma_v^2 = \frac{1}{N_s}\mathbf{v}^t\mathbf{v} \leq 1 \\ 2\mathbf{b} - 1 = \text{sign}\left(R_{\tilde{\mathcal{D}}}\tilde{\mathcal{D}}^t\left(\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}}\right)\right), \end{cases} \tag{10.15}$$

  où $\mathbf{b}$ est le K-uplet à transmettre, $R_{\tilde{\mathcal{D}}}$ est la matrice d'autocorrélation du dictionnaire de réception $\tilde{\mathcal{D}}$ et $\tilde{\mathbf{p}}$ est un bruit de canal.

- dans le cas de l'égaliseur de Larbi,

$$\begin{cases} \sigma_v^2 = \frac{1}{N_s}\mathbf{v}^t\mathbf{v} \leq 1 \\ 2\mathbf{b} - 1 = \text{sign}\left(R_{\mathcal{D}}\mathcal{D}^t\left(\tilde{\mathbf{x}} + \alpha\tilde{\mathbf{v}} + \tilde{\mathbf{p}}\right)\right) \end{cases} \tag{10.16}$$

où $R_{\mathcal{D}}$ est la matrice d'autocorrélation du dictionnaire $\mathcal{D}$.

La stratégie d'insertion informée se propose de rechercher le signal modulé $\mathbf{v}$ sous la forme d'une combinaison linéaire des vecteurs du dictionnaire d'émission $\mathcal{D}$:

$$\mathbf{v} = \sum_{m=0}^{M-1} \gamma_m \left(2b[m] - 1\right) \mathbf{d}_m, \tag{10.17}$$

dont les coefficients γ_m sont spécifiquement choisis en fonction du signal audio pour maximiser la robustesse de la transmission aux perturbations.

Le calcul du dictionnaire $\tilde{\mathcal{D}}$ et l'expression de $\mathbf{v}$ sous la forme d'une combinaison linéaire des vecteurs du dictionnaire établissent un lien efficace entre le signal modulé $\mathbf{v}$ et sa version filtrée $\tilde{\mathbf{v}}$, donné par :

$$\begin{aligned} \mathbf{v} &= \sum_{m=0}^{M-1} \gamma_m \left(2\mathbf{b}[m] - 1\right) \mathbf{d}_m = \mathcal{D}(\gamma \otimes (2\mathbf{b} - 1)) \\ &\Longleftrightarrow \tilde{\mathbf{v}} = \sum_{m=0}^{M-1} \gamma_m \left(2\mathbf{b}[m] - 1\right) \tilde{\mathbf{d}}_m = \tilde{\mathcal{D}}(\gamma \otimes (2\mathbf{b} - 1)). \end{aligned} \tag{10.18}$$

La recherche du signal modulé $\mathbf{v}$ adapté et de sa version filtrée $\tilde{\mathbf{v}}$ peut donc être effectuée conjointement.

L'étude du modèle informé nous a permis de résoudre un problème équivalent, celui de trouver $\mathbf{v} = \mathcal{D}((2\mathbf{b} - 1) \otimes \gamma)$, tel que :

$$\begin{cases} \sigma_v^2 = \frac{1}{N_s}\mathbf{v}^t\mathbf{v} \leq 1 \\ 2\mathbf{b} - 1 = sign\left(R_{\mathcal{D}}^{-1}\mathcal{D}^t\left(\mathbf{x} + \alpha\mathbf{v} + \mathbf{p}\right)\right) \end{cases} \tag{10.19}$$

en introduisant un paramètre de robustesse aux perturbations τ_p, qui permet d'aboutir à un problème d'optimisation sur la recherche du vecteur d'amplitude γ :

$$\begin{cases} \gamma = \arg\max_\mu \tau_p^2(\mu) \\ \tau_p^2(\mu) = \min\left(\alpha\mu + (2\mathbf{b} - 1) \otimes \mathcal{D}^\dagger\mathbf{x}\right) \\ \mu^t\mu = 1 \end{cases} \tag{10.20}$$

Ce problème est résolu par un algorithme itératif qui consiste à incrémenter petit à petit les composantes de γ jusqu'à atteindre la puissance unité du signal modulé en sélectionnant à chaque étape la composante pour laquelle la transmission est la plus défavorable. Cette composante est celle qui minimise $\left(\mu + (2\mathbf{b} - 1) \otimes \tilde{\mathcal{D}}^\dagger\mathbf{x}\right)$ ou de manière équivalente celle qui minimise $(2\mathbf{b} - 1) \otimes \left(R_{\mathcal{D}}^{-1}\mathcal{D}^t\left(\mathbf{x} + \alpha\mathbf{v} + \mathbf{p}\right)\right)$.

L'application de la stratégie d'insertion au système de tatouage en boucle fermée s'effectue par analogie avec le modèle de la chaîne :

- dans le cas du filtre blanchissant, le problème d'optimisation se réécrit sous la forme :

$$\begin{cases} \gamma = \arg\max_\mu \tau_p^2(\mu) \\ \tau_p^2(\mu) = \min\left(\alpha\mu + (2\mathbf{b} - 1) \otimes \tilde{\mathcal{D}}^\dagger\tilde{\mathbf{x}}\right) \\ \mu^t\mu = 1 \end{cases} \tag{10.21}$$

L'algorithme mis en œuvre pour sa résolution est donc identique au cas précédent à condition que la composante de γ à chaque itération soit sélectionner pour minimiser $\Big(\alpha\mu + (2\mathbf{b}-1)\otimes\tilde{\mathcal{D}}^{\dagger}\tilde{\mathbf{x}}\Big)$.

- dans le cas de l'égaliseur, la formalisation du problème est plus complexe. En effet, le dictionnaire $\tilde{\mathcal{D}}$ utilisé pour calculer le tatouage $\tilde{t}$ en sortie de l'égaliseur est différent du dictionnaire $\mathcal{D}$ exploité pour la réception. Le problème ne peut donc être formalisé (et simplifié) sous une forme aussi simple que dans le cas du filtre blanchissant. Il s'écrit :

$$\begin{cases} \gamma = \arg\max_{\mu} \tau_p^2(\mu) \\ \tau_p^2(\mu) = \min\Big[\alpha(2\mathbf{b}-1)\otimes\Big(R_{\mathcal{D}}^{-1}\mathcal{D}^t\tilde{\mathcal{D}}\left(\gamma\otimes(2\mathbf{b}-1)\right)\Big) + \mathcal{D}^{\dagger}\tilde{\mathbf{x}}\Big] \\ \mu^t\mu = 1 \end{cases} \tag{10.22}$$

Pourtant, malgré ce formalisme plus complexe, l'algorithme mis en œuvre est identique : il consiste à incrémenter itération après itération les composantes de γ en sélectionnant à chaque étape celle qui minimise $\Big[\alpha(2\mathbf{b}-1)\otimes\Big(R_{\mathcal{D}}^{-1}\mathcal{D}^t\tilde{\mathcal{D}}\left(\gamma\otimes(2\mathbf{b}-1)\right)\Big) + \mathcal{D}^{\dagger}\tilde{\mathbf{x}}\Big]$.

10.3.2 Spécificités de l'égaliseur de Wiener

Dans le cas de l'égaliseur de Larbi (contrairement au cas du filtre blanchissant), le filtre de Wiener est adapté au signal $\mathbf{z}$ somme de la composante audio blanchie par $\dfrac{1}{H(f)}$ et du signal modulé. Le filtre de Wiener est donc dépendant du vecteur du dictionnaire choisi pour la modulation et influence le signal audio $\tilde{\mathbf{x}}$ utilisé pour sélectionner le signal modulé optimal. Ce signal n'étant déterminé au regard du résultat du filtrage, le filtre de Wiener est donc calculé à partir du signal modulé, qui serait construit si l'on utilisait la stratégie d'insertion non informée.

10.3.3 Performances

Cette section propose d'établir les performances de la stratégie d'insertion informée adaptée aux modulations CDMA, au travers des quatre critères définis par le protocole expérimental : l'inaudibilité, le TEB pour un canal sans perturbation, la robustesse aux perturbations et le coût en temps de calcul. Le dictionnaire sera constitué de $M = K = 4$ vecteurs, permettant un encodage de 4 bits par symbole. Les performances du système en boucle fermée seront comparées aux performances du système non informé de référence établies chapitre 7.

Inaudibilité

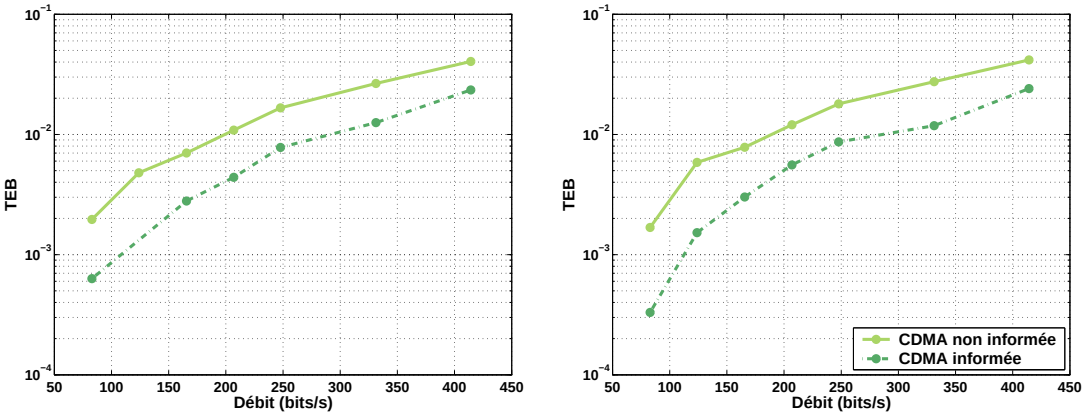
La table 10.1 présente les valeurs moyennes de l'ODG et les écart-types obtenues par tatouage des 20 signaux du corpus à 206 bit/s pour les trois systèmes considérés. Ces mesures montrent que l'introduction de la stratégie d'insertion informée quel que soit le récepteur utilisé dans la boucle fermée ne modifie pas la transparence du tatouage qui reste audible mais non gênante.

TEB pour un canal sans perturbations

La figure 10.4 présente les TEBs obtenus par le système informé pour les deux types de récepteurs (filtre blanchissant dans le cas de la figure (a) et égaliseur de Larbi dans le cas de la figure (b)) en fonction du débit lorsque le canal ne subit pas de perturbations. Les performances du système non informé sont également rappelées. Ces courbes mettent en évidence

Système	Non informé	Informé	
Récepteur		Filtre blanchissant	Égaliseur de Larbi
ODG	$-1,0\ (0,07)$	$-1,0\ (0,07)$	$-1,0\ (0,07)$

Table 10.1 – Valeurs moyennes de l'ODG et écart-type (entre parenthèses) pour le système de tatouage avec modulation CDMA non informé et informé avec les deux types de récepteur.

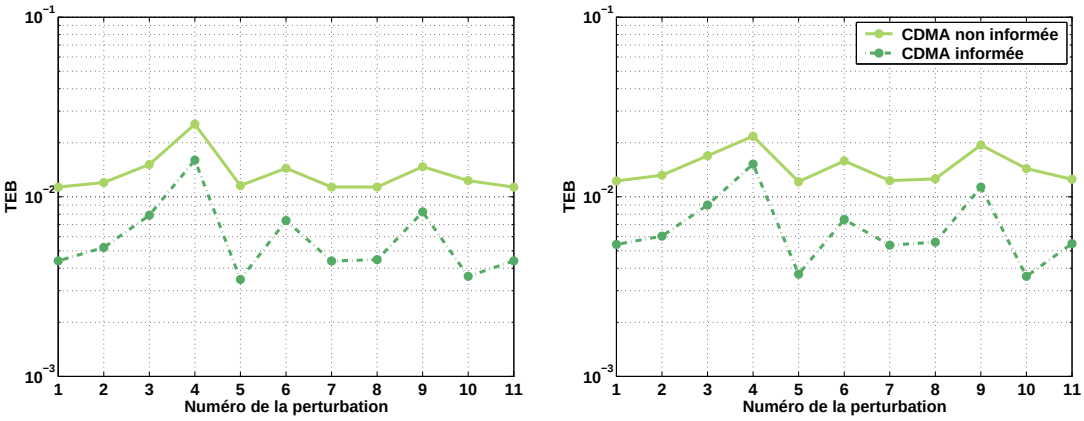


(a) (b)

Figure 10.4 – Comparaison des performances du système de tatouage CDMA non informé et de la stratégie d'insertion informée pour les deux récepteurs : (a) le filtre blanchissant et (b) l'égaliseur de Larbi. Le canal ne subit pas de perturbations. Le dictionnaire contient $M = K = 4$ vecteurs orthogonaux.

l'amélioration des TEBs du système grâce à l'utilisation de la stratégie d'insertion informée et ce quel que soit le récepteur : en effet, les TEBs sont environ divisés par 2. A fiabilité de transmission égale, le système informé permet donc un accroissement du débit de l'ordre de 50 bit/s pour de TEBs de 2.10^{-3} et de l'ordre de 100 bit/s pour des TEBs de 10^{-2}.

En comparant ces courbes aux performances des stratégies d'insertion proposées pour les modulations M-aire figure 9.14, nous pouvons conclure que, pour un même paramètre d'encodage K, les stratégies informées pour les modulations CDMA sont moins efficaces que la stratégie informée finale proposée pour les modulations M-aire. En fait, l'amélioration apportée pour modulation CDMA est comparable à celle apportée par la stratégie informée M-aire avec choix d'un tatouage inaudible et de robustesse maximale aux perturbations mais sans mise en œuvre du dictionnaire structuré : en effet, l'algorithme choisissant le tatouage est assez similaire, puisqu'il consiste à diriger pas à pas le tatouage vers la région de détection. Par contre, la stratégie CDMA, contrairement à la stratégie M-aire, ne peut bénéficier des apports d'un dictionnaire structuré : en effet, pour mettre en place un tel dictionnaire, il faut que la procédure de sélection du vecteur du dictionnaire structuré $\mathbf{d}_k^{opt}$ adapté à la transmission de l'information adopte le même formalisme que le procédé de détection. Cette procédure s'attache à rechercher le vecteur de plus grande corrélation avec le signal audio alors que la détection pour les modulations CDMA tient au signe de cette corrélation ; elle n'est donc pas compatible avec la détection de l'information pour les modulations CDMA.



(a) (b)

Figure 10.5 – TEBs du système CDMA informé et du système non informé à un débit fixe de 206 bit/s, en fonction des différentes perturbations de canal (ici numérotées), lorsque le récepteur est : (a) le filtre blanchissant, (b) l'égaliseur de Larbi.

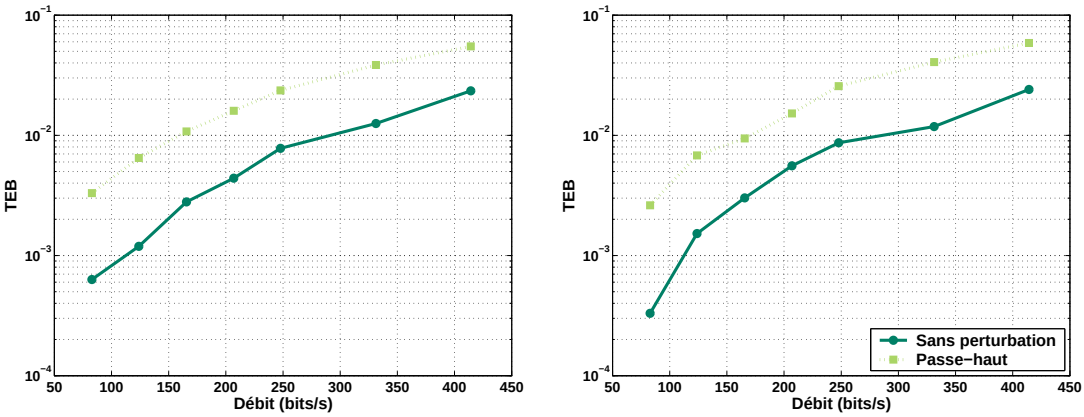
Robustesse aux perturbations

La robustesse aux perturbations est maintenant étudiée. Nous présentons figure 10.5 les TEBs obtenus par la stratégie d'insertion informé pour les deux récepteurs à 206 bit/s en fonction de la perturbation considérée. Cette perturbation est présentée sous la forme d'un numéro, avec la concordance établie table 10.2. Cette figure rappelle également les performances en matière de robustesse de la stratégie d'insertion non informée.

Numéro	Perturbations	Numéro	Perturbations
1	Sans perturbations	**7**	Sous échantillonnage (22050 kHz)
2	Compression MPEG 96 kbit/s	**8**	Quantification (12 bits)
3	Compression MPEG 64 kbit/s	**9**	Ajout d'un bruit blanc (30 dB)
4	Filtrage passe-haut* (250 Hz)	**10**	Amplification (réduction de 50%)
5	Filtrage passe-bas* (9 kHz)	**11**	Compression de dynamique*
6	Ajout d'écho		

Table 10.2 – Liste des perturbations et de leur numéro associé. Les perturbations indiquées par une * sont réalisées par le Stirmark Audio.

Ces courbes font état une fois encore de la robustesse du système de tatouage informé aux perturbations. Les TEBs du système informé quel que soit le récepteur considéré varient relativement peu en fonction de la perturbation, à l'exception de la perturbation par filtrage passe-haut. Cette perturbation est une fois encore la plus dégradante et entraîne une augmentation des TEBs supérieure à $3,8$. La figure 10.6 présente de fait les TEBs du système CDMA informé en fonction du débit pour ce filtrage passe-haut et les compare aux TEBs du système sans perturbation. Compte tenu de la dégradation introduite, une transmission fiable et robuste de l'information pour un TEB de 2.10^{-3} peut être obtenue à environ 120 bit/s contre 75 bit/s pour le système non informé.



(a) (b)

Figure 10.6 – TEBs en fonction du débit obtenus par le système CDMA informé lorsque le canal ne subit pas de dégradation et lorsqu'il est soumis à un filtre passe-haut. Le récepteur utilisé est : (a) le filtre blanchissant et (b) l'égaliseur de Larbi

	Émetteur	**Récepteur**
Non informé	78%	15,7%
Informé avec filtre blanchissant	220%	47%
Informé avec égaliseur de Larbi	504%	15,7%

Table 10.3 – Coût en temps de calcul par rapport au temps-réel des différentes configurations d'émission et de réception.

Coût en temps de calcul

La table 10.3 présente à titre indicatif le coût en temps de calcul de l'émetteur et du récepteur des deux systèmes informés basés sur une modulation CDMA et le compare au coût du système non informé. Comme dans le cas des stratégies d'insertion informées pour les modulations M-aire, l'introduction de la copie locale du récepteur à l'émetteur et le choix itératif du tatouage adapté conduit à une augmentation du coût de l'émetteur. La stratégie basée sur le filtre blanchissant est moins coûteuse que celle exploitant l'égaliseur, puisque cette dernière requière un plus grand nombre d'opérations du filtrage (les filtres étant dépendant du signal de tatouage choisi). Le coût de la réception reste inchangée puisque qu'aucune modification du récepteur ni du dictionnaire n'a été effectuée. La faisabilité d'une telle stratégie d'insertion en temps-réel est donc envisageable si le dictionnaire contient un faible nombre de vecteurs.

Conclusion

Cette section s'est concentrée sur l'étude des stratégies d'insertion informées basées sur les modulations CDMA. Un signal CDMA est construit comme la combinaison des vecteurs du dictionnaire dont les coefficients dépendent de l'information binaire à émettre et d'un facteur d'amplitude. La stratégie informée que nous proposons consiste à choisir le facteur d'amplitude spécifiquement adapté au signal audio pour garantir la transmission inaudible et correcte de

l'information. Elle exploite, pour se faire, la copie locale du récepteur à l'émetteur introduite dans le système de tatouage. Cette structure en boucle fermée permet, grâce à un algorithme itératif, de faire converger pas à pas le tatouage vers la région de détection de l'information sous contrainte d'inaudibilité. Les performances de cette stratégie ont été évaluées sur un modèle de la chaîne de tatouage puis sur la structure du système en boucle fermée définie pour ces deux récepteurs. Sa mise en œuvre sur des signaux réels établit son efficacité par rapport au système non informé : la transparence du tatouage est correctement maintenue, le TEB est au moins divisé par 2 et la robustesse du tatouage est équivalente au cas non informé. Ainsi, grâce à la stratégie informée que nous proposons, une transmission de TEB 10^{-3} peut être obtenue à 120 bit/s pour un canal sans perturbations (contre environ 80 bit/s pour le système non informé). Cette stratégie montre néanmoins des performances limitées et inférieures à celles obtenues par les stratégies informées adaptées aux modulations M-aire.

Conclusion

Dans cette partie, nous avons proposé de faire tendre le système de tatouage développé dans la deuxième partie vers un système informé, c'est à dire exploitant la connaissance *a priori* du signal audio à l'émetteur. Dans cet objectif, le système est modifié pour adopter une structure en boucle fermée : une copie locale du récepteur est introduite à l'émetteur. Dans la seconde partie, nous avions proposé un récepteur exploitant un filtre blanchissant et un dictionnaire de réception, version filtrée du dictionnaire d'émission. Les performances de ce récepteur avaient été comparées à un récepteur de l'état de l'art basé sur une procédure d'égalisation. Ces deux récepteurs ont donc été considérés dans le système en boucle fermée donnant lieu à deux schémas d'émetteur informé.

La copie locale de ces récepteurs à l'émetteur nous a permis d'établir les conditions à satisfaire par le tatouage $\mathbf{t}$ pour garantir une transmission inaudible, correcte et robuste de l'information cachée face à un bruit de canal $\mathbf{p}$ supposé additif et non désynchronisant. L'inaudibilité relève d'une contrainte de puissance, imposée par le filtre de mise en forme perceptuelle et le gain α qui régissent la mise en forme spectrale du signal de tatouage $\mathbf{t}$. Les contraintes de détection correcte et robuste sont spécifiques à la modulation utilisée (dans notre cas, une modulation M-aire ou une modulation CDMA) et au récepteur considéré. Elles peuvent être exprimées à l'aide d'une estimation, obtenue grâce à la copie locale du récepteur, des signaux intervenant dans le procédé de détection. La contrainte de détection correcte et robuste prend alors la forme d'un ensemble d'inégalités traduisant le résultat correct du démodulateur par corrélation dans le cas des modulations M-aire et du démodulateur par décorrélation dans le cas des modulations CDMA. L'émetteur informé relève alors d'une stratégie : celle de trouver le tatouage qui soit solution des inégalités de détection pour un bruit de perturbation de puissance maximale sous la contrainte de puissance traduisant l'inaudibilité. Dans cet objectif, le tatouage recherché est issu d'une combinaison linéaire des vecteurs du dictionnaire ; ce formalisme nous a permis de spécifier très simplement l'expression du tatouage au travers des différentes étapes de la chaîne (modulation, mise en forme spectrale et réception). Pour résoudre ce problème, nous avons ensuite considéré séparément les modulations M-aire et les modulations CDMA.

Dans le cas de modulation M-aire, nous avons proposé deux approches :

- la première approche consiste à modifier le dictionnaire d'émission pour lui conférer une structure similaire à celle du dictionnaire de Costa[Cos83]. Le dictionnaire est scindé en sous-dictionnaires, contenant chacun un ensemble de vecteurs susceptibles de transmettre l'information. Pour chaque insertion, le vecteur maximisant les chances de détecter correctement l'information insérée est sélectionné. L'état de l'Art avait déjà établi l'efficacité de cette approche (en terme de capacité de transmission et sur des mesures expérimentales pour le tatouage d'images). Dans cette étude, nous en avons établi les performances théoriques, en terme de probabilité d'erreur, et nous avons mesuré ses performances pour le système de tatouage audio sur des signaux réels.

- la seconde approche consiste à choisir le tatouage, issu d'une combinaison linéaire des vecteurs du dictionnaire, solution des inégalités de détection robuste en présence du bruit de canal $\mathbf{p}$ sous la contrainte d'inaudibilité. Cette approche conduit à deux problèmes d'optimisation dépendants de la formalisation de la robustesse. Dans le premier, le bruit de canal est modélisé par un paramètre dit de robustesse ; le tatouage est alors choisi de sorte à maximiser ce paramètre. Dans le second, le bruit est modélisé par un BBAG, permettant d'exprimer la probabilité d'erreur de transmission P_e face au bruit. Le tatouage est alors choisi de sorte à atteindre une probabilité d'erreur fixée pour une puissance de bruit maximale. Les performances de ces deux stratégies ont été évaluées sur un modèle théorique de la chaîne de tatouage, pour ne sélectionner que la plus efficace (la première) et l'appliquer à l'évaluation des performances du système de tatouage en boucle fermée sur des signaux réels.

Les deux approches ont finalement été combinées en une stratégie unifiée dont les performances finales ont été évaluées sur des signaux réels.

Dans le cas des modulations CDMA, où le tatouage permet la transmission simultanée de plusieurs bits, nous avons proposé de choisir spécifiquement l'énergie dédiée à la transmission de chaque bit pour garantir la robustesse et l'inaudibilité de la transmission. La robustesse est ici formalisée par un paramètre de bruit à maximiser. Le choix de l'énergie adaptée est établi sous la forme d'un problème d'optimisation que nous avons résolu par un algorithme similaire à celui sélectionné pour les modulations M-aire. Les performances de la stratégie d'insertion informée ont été évaluées sur un modèle de la chaîne de tatouage et sur le système de tatouage en boucle fermée avec des signaux audio réels.

L'évaluation des performances a établi que les stratégies d'insertion informée, qu'elles soient adaptées aux modulations CDMA ou aux modulations M-aire, garantissent une transparence très correcte du tatouage (un tatouage audible mais non gênant) et une robustesse de la transmission tout à fait comparable aux systèmes non informés : une augmentation des TEBs par 2 est à prévoir pour la perturbation la plus dégradante en comparaison du cas sans perturbation. Les TEBs évalués pour un canal sans perturbation montrent une très nette amélioration des performances par rapport au cas non informé. En la matière, la stratégie adaptée aux modulations M-aire est plus efficace que celle construite pour les modulations CDMA. Elle permet d'obtenir une transmission de TEB 10^{-3} à un débit de 170 bit/s grâce au récepteur par égalisation (contre 80 bit/s dans le cas non informé) lorsque le canal n'est pas dégradé. Ce débit doit être diminué aux alentours de 75 bit/s si des perturbations non désynchronisantes sont introduites.

L'introduction des stratégies informées dans le système de tatouage additif par étalement de spectre a donc permis une amélioration significative de la fiabilité de transmission en fonction du débit tout en conservant une transparence du tatouage et une robustesse de la transmission aux perturbations non désynchronisantes très acceptables. Il convient maintenant d'intégrer au système un mécanisme de synchronisation adapté permettant de garantir sa robustesse aux opérations désynchronisantes.

Quatrième partie

Synchronisation

Chapitre 11

Mécanisme de synchronisation pour des opérations désynchronisantes à forte dérive

L'une des catégories de perturbations les plus dégradantes pour un système de tatouage est sans aucun doute celle des perturbations désynchronisantes. Plusieurs types de désynchronisation peuvent être répertoriés, dépendant de l'application à laquelle se destine le système de tatouage. Ces désynchronisations peuvent aller de l'introduction d'un simple retard lors d'un filtrage à la dilation ou la contraction de la durée du signal audio lors, par exemple, de conversions analogique-numérique ou d'une opération de "*time stretching*" [Zöl03]. Elles nécessitent donc la mise en place d'un système de synchronisation spécifique qui peut être dissocié de la technique d'insertion de données.

Ces types de perturbations peuvent être modélisés par un rééchantillonnage à la fréquence $\hat{F}_e$ du signal audio tatoué $y(n)$, initialement échantillonné à la fréquence F_e. Ce rééchantillonnage introduit une dérive entre l'échelle des temps de l'émission et l'échelle des temps de réception. Il induit une modification de la localisation des symboles transmis mais également une modification de leur durée. Nous définirons la dérive d comme l'erreur relative de la fréquence de réception par rapport à celle de l'émission :

$$d = \frac{\hat{F}_e - F_e}{F_e} \tag{11.1}$$

Cette dérive peut prendre des valeurs faibles, typiquement de l'ordre de $0,1\%$ si l'on souhaite utiliser le canal audio pour transmettre de l'information entre deux PCs par exemple, ou nettement plus élevées, de l'ordre de quelques pour cent, dans des applications de type radiodiffusion. En pratique ces dilations ou contractions du signal audio tatoué peuvent atteindre jusqu'à 10% de la durée initiale du signal audio [KM03]. La dérive peut être constante dans le cas d'un simple rééchantillonnage après conversion analogique du signal audio. Mais elle peut également présenter de fortes discontinuités en radiodiffusion. En effet, une émission radio peut être entrecoupée de séquences sonores, préalablement tatouées et devant être diffusées dans un temps imparti. Le "*time stretching*" permet alors de modifier pour l'occasion la durée de la séquence de sorte à satisfaire la contrainte de temps de diffusion.

L'état de l'art en matière de synchronisation des systèmes de tatouage audio, que nous avons présenté section 3.4, établit deux grandes familles de systèmes de synchronisation. La

première regroupe les systèmes exploitant des formes d'onde étalées spectralement et leur propriété d'autocorrélation. Ces formes d'onde sont utilisées soit en guise de pattern de synchronisation, soit pour la modulation de l'information à tatouer. Elles permettent d'estimer l'opération désynchronisante, en évaluant localement la dérive entre l'émission et la réception et d'inverser l'opération avant de procéder à la détection du tatouage. Souvent mis en œuvre pour des opérations désynchronisantes illicites (cas d'un pirate qui ajouterait ou supprimerait aléatoirement des échantillons du signal audio pour empêcher la détection du tatouage-marque), ces systèmes nécessitent un temps de calcul relativement important qui ne permet pas d'envisager un traitement temps-réel du récepteur. La seconde famille de systèmes regroupe les mécanismes de synchronisation qui extraient du signal audio des points d'intérêts, qui définissent une cartographie du signal. L'extraction de cette "carte" à la réception permet là encore d'estimer et d'inverser l'opération désynchronisante pour procéder ensuite à la détection de l'information cachée. Malheureusement, ces points d'intérêt étant souvent choisis en fonction des caractéristiques du signal, les systèmes de tatouage qui en découlent induisent un débit de transmission variable. Un mécanisme de synchronisation adapté aux désynchronisations licites, quasi temps-réel et induisant un débit de transmission constant fait donc défaut à la littérature.

Dans cette partie nous nous intéressons à la construction d'un tel mécanisme de synchronisation. Ce mécanisme de synchronisation est dédié aux perturbations désynchronisantes pouvant être modélisées par un rééchantillonnage du signal audio tatoué dont la dérive peut prendre des valeurs importantes allant jusqu'à $\pm 2\%$ [1] mais ne présente pas de variations [2]. Il ne s'intéresse qu'aux systèmes basés sur des modulations M-aires et non aux modulations CDMA. En effet, les études précédentes ont établi l'efficacité des modulations M-aire informées sur les modulations CDMA.

11.1 Modélisation de la désynchronisation

Pour permettre la détection correcte de l'information cachée, l'objectif du mécanisme de synchronisation est d'établir la correspondance entre l'échelle des temps à l'émission et l'échelle des temps à la réception. Pour établir une distinction claire entre ces deux échelles, nous noterons respectivement n et $\hat{n}$ les échantillons à l'émission et à la réception. Cette correspondance peut être établie par une modélisation explicite des opérations désynchronisantes.

Les perturbations désynchronisantes licites que nous considérons dans cette thèse sont de deux natures :

- elles peuvent provenir de l'ajout d'un retard entre le signal audio tatoué à l'émission et à la réception, notée $\hat{y}$. La modification de l'échelle des temps est alors caractérisée par le facteur de retard $\hat{n}_0$.

- elles peuvent être liées à la dilation ou la contraction du signal audio tatoué lors de la transmission du signal sous format analogique ou d'opérations de "*time stretching*". Dans ce contexte, la désynchronisation peut être modélisée par un rééchantillonnage du signal audio tatoué $y(n)$ d'une fréquence d'échantillonnage F_e à l'émission à une fréquence $\hat{F}_e$

1. Cette limite sur la dérive est plus faible que celle de 10% proposé dans [KM03] mais représente déjà une différence d'environ 1000 échantillons par seconde entre les signaux à l'émission et à la réception, s'ils sont échantillonnés à $44,1$ kHz.

2. La concaténation de deux signaux dilatés ou comprimés avec deux facteurs de dérives différents ne sera notamment pas considérées.

à la réception $\hat{y}(\hat{n})$. La modification induite de l'échelle des temps peut alors être caractérisée par un facteur de dérive d, exprimé comme l'écart relatif entre les fréquences d'échantillonnage :

$$d = \frac{\hat{F}_e}{F_e} - 1. \tag{11.2}$$

Si $d > 0$, la perturbation correspondra à une dilation ; si $d < 0$, il s'agira d'une contraction.

11.1.1 Influence du retard

L'introduction d'un retard (s'il n'est pas combiné à un rééchantillonnage du signal ou à une interpolation) conduit simplement à translater l'échelle de réception par rapport à l'échelle d'émission d'un facteur $\hat{n}_0$. Ces deux échelles sont donc liées par la relation :

$$\hat{n} + \hat{n}_0 = n \tag{11.3}$$

de sorte que $\hat{y}(\hat{n}) = y(n - \hat{n}_0)$. Ce retard introduit donc une modification de la localisation des symboles dans le signal audio tatoué mais pas de modification de leurs durées N_s. La seule estimation du retard $\hat{n}_0$ permet donc de compenser la désynchronisation.

11.1.2 Influence de la dérive

Considérons le signal audio tatoué $y(n)$ échantillonné à la fréquence F_e et sa version rééchantillonnée à la réception $\hat{y}(\hat{n})$ à une fréquence d'échantillonnage $\hat{F}_e$. L'expression de $\hat{y}(\hat{n})$, en utilisant la formule de Shannon pour la conversion analogique du signal $y(n)$, est la suivante :

$$\hat{y}(\hat{n}) = \sum_{n=-\infty}^{\infty} y(n) \sin_c \left(F_e \left(\frac{\hat{n}}{\hat{F}_e} - \frac{n}{F_e} \right) \right). \tag{11.4}$$

Établir la relation existant entre l'échelle des temps de réception $\hat{n}$ et celle d'émission n consiste à trouver pour chaque échantillon $\hat{y}(\hat{n})$ du signal audio tatoué reçu l'échantillon du signal audio tatoué $y(n)$ le plus corrélé avec $\hat{y}(n)$. Cette corrélation dépend de la contribution du sinus cardinal, qui est maximale lorsque $\hat{n} = \frac{\hat{F}_e}{F_e} n$. $\frac{\hat{F}_e}{F_e}$ n'étant pas nécessairement entier, n ne l'est pas nécessairement non plus ; la plus grande contribution du sinus cardinal est donc apportée par : $n(\hat{n}) = \left[\frac{F_e}{\hat{F}_e} \hat{n} \right]$ où $[x]$ caractérise l'arrondi de x. En introduisant l'écart $\delta(\hat{n}) = n(\hat{n}) - \frac{F_e}{\hat{F}_e} \hat{n}$, nous obtenons l'équation caractérisant la dérive temporelle, c'est à dire l'expression de l'indice n de l'échelle d'émission en fonction de l'indice de l'échelle de réception :

$$n(\hat{n}) = \left[\frac{F_e}{\hat{F}_e} \hat{n} \right] + \delta(\hat{n}). \tag{11.5}$$

Cette équation met en évidence deux phénomènes : d'une part, l'opération désynchronisante introduit une modification de la localisation des symboles mais également une modification de leurs durées : en effet, deux symboles distants de N_s échantillons à l'émission sont désormais distants de $\hat{N}_s = \frac{\hat{F}_e}{F_e} N_s$ dans l'échelle de réception.

L'équation caractérisant la dérive (11.5) permet également de réécrire le signal audio désynchronisé sous la forme interpolée suivante :

$$\begin{aligned}\hat{y}(\hat{n}) &= \sum_{i=-\infty}^{\infty} y\left(n(\hat{n})+i\right)\sin_c\left(i+\delta(\hat{n})\right) \\ &= \sin_c\left(\delta(\hat{n})\right)y(n(\hat{n})) + \sum_{i\neq 0} y(n(\hat{n})+i)\sin_c\left(i+\delta(\hat{n})\right). \end{aligned} \tag{11.6}$$

En plus de la désynchronisation introduite par la modification de l'échelle des temps, le rééchantillonnage du signal audio tatoué introduit une modification de l'amplitude du signal (dont l'amplitude dépend du temps) et l'ajout d'un bruit coloré, comme cela a également été montré dans [PLG03, ZBD04].

11.1.3 Objectif du mécanisme de synchronisation

En considérant une opération désynchronisante qui introduit à la fois un retard et un rééchantillonnage du signal audio tatoué, le signal obtenu à l'entrée du récepteur peut finalement s'exprimer sous la forme :

$$\hat{y}(\hat{n}+\hat{n}_0) = \sum_{q=-\infty}^{\infty} y\left(n(\hat{n})+q\right)\sin_c\left(q+\delta(\hat{n})\right) \tag{11.7}$$

La modification de l'échelle des temps peut être entièrement déterminée par la donnée du retard $\hat{n}_0$ et de la fréquence d'échantillonnage à la réception $\hat{F}_e$, qui permet de déterminer aussi bien le début de chaque symbole que leur durée à la réception. Le mécanisme de synchronisation à mettre en œuvre doit donc permettre d'estimer ce retard $\hat{n}_0$ et cette fréquence d'échantillonnage. En utilisant l'équation caractérisant la dérive (11.5), la localisation des symboles dans le signal reçu pourrait alors être déduite. Un mécanisme de synchronisation très simple consistant à introduire une séquence d'apprentissage en amont du tatouage permet de réaliser aisément le calcul des paramètres de la désynchronisation.

Néanmoins, une erreur d'estimation sur la fréquence d'échantillonnage (même faible) est à prendre en compte. Dans ce cas, la localisation des symboles (si elle est calculée d'après (11.5) est erronée et la détection n'est plus synchronisée. Un mécanisme de synchronisation locale, incorporant la valeur de $\hat{F}_e$, et permettant la synchronisation du système sur chaque temps symbole doit donc être envisagé.

11.2 Mécanisme de synchronisation

Pour réaliser l'estimation des paramètres de l'opération désynchronisante et la synchronisation locale sur les temps symbole du récepteur, le mécanisme de synchronisation proposé s'inspire des mécanismes utilisés en communication numérique par étalement de spectre [Pro01]. Son schéma de principe est présenté figure 11.1. Il se découpe en trois phases :

- la première phase est une ***phase d'initialisation***. Elle a pour but d'estimer le retard à l'origine $\hat{n}_0$ et la fréquence $\hat{F}_e$ avant l'envoi de l'information binaire à tatouer.

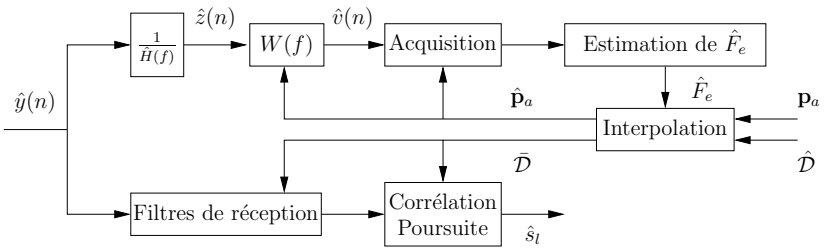


Figure 11.1 – Schéma de principe du mécanisme de synchronisation pour les phases d'acquisition et de poursuite.

- la seconde phase, dite ***d'acquisition***, réalise la synchronisation sur des points de référence, qui créent une cartographie du signal audio à intervalle de temps régulier (indépendant du signal audio) et permettent d'actualiser la valeur de $\hat{F}_e$. Cette actualisation permet d'annuler progressivement des erreurs d'estimation sur $\hat{F}_e$ voire de prendre en compte d'éventuelles variations (faibles) de la dérive.

- la troisième phase, dite ***de poursuite***, permet de réaliser la synchronisation locale. Elle permet de synchroniser le récepteur sur chaque temps symbole et de procéder à la détection des symboles localisés.

Notre système de tatouage effectuant l'insertion du tatouage à l'aide de formes d'onde étalées puis mises en forme spectralement, nous nous sommes naturellement dirigés vers l'utilisation de patterns de synchronisation. Ces patterns, formes d'onde étalées spectralement et spécifiquement dédiées à la synchronisation, sont insérées tout au long du message en suivant la même stratégie d'insertion que celle du système de tatouage. Leur insertion induit une structure particulière du signal modulé et permet la réalisation des trois phases du mécanisme de synchronisation.

11.2.1 Structure du signal modulé

L'introduction du mécanisme de synchronisation nécessite de modifier la structure du signal modulé. Précédemment ce signal était construit comme la concaténation sur chaque temps symbole de formes d'onde modulant la suite de symboles à émettre. Pour les besoins du mécanisme de synchronisation, ce signal est entrecoupé de patterns, formes d'onde étalées spectralement, dédiées à la synchronisation, conformément au schéma donné figure 11.2. Ces patterns sont supposées connues du récepteur.

Deux patterns de synchronisation sont considérées. La première, notée $\mathbf{p}_i$, de durée N_i, est exploitée par la phase d'initialisation. Cette pattern est donc insérée M_i fois en amont du signal modulé à intervalle de $N_i + N_z$ échantillons. Les N_z échantillons séparant deux patterns d'initialisations sont constitués de 0. La seconde pattern notée $\mathbf{p}_a$ est insérée pour les besoins de la phase d'acquisition. Elle nécessite de découper le signal modulé (et la séquence de symboles $\{s_l\}$ à émettre) sous la forme de messages de longueur N_m. Chaque message est constitué une pattern de synchronisation $\mathbf{p}_a$ de durée N_a servant de points de référence et du signal modulé associé à une sous-séquence de symboles. Cette sous-séquence est une suite de L_m symboles $\{s_l^m\}_{l\in[\![0,L_m-1]\!]}$ extraits de la séquence des symboles à émettre, où m indique le numéro du

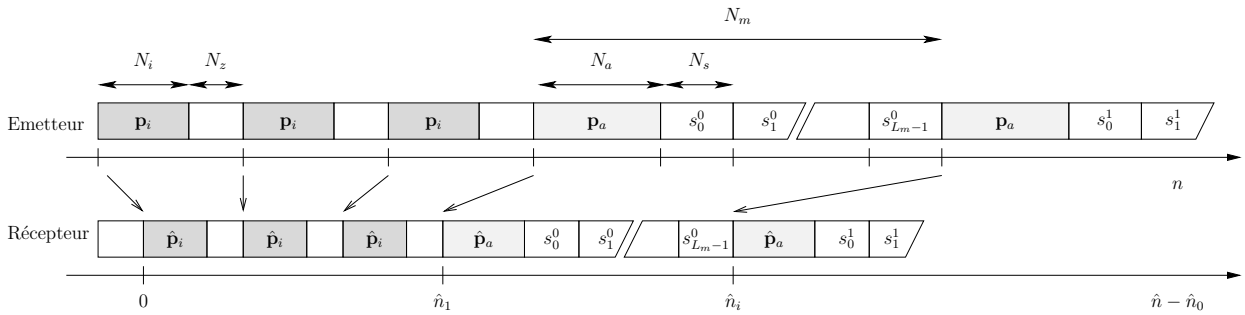


Figure 11.2 – Structure du signal modulé pour le mécanisme de synchronisation à l'émission et à la réception (après dilation par un facteur 2/3).

message.

Les patterns sont insérées au signal audio selon le même procédé que le signal modulé. Elles sont donc choisies étalées spectralement dans la bande de fréquence $[0, F_c]$ et de puissance unité. Pour leur insertion, elles sont filtrées par le filtre de mise en forme spectrale $H(f)$, pondérées par un facteur de gain α choisi spécifiquement en fonction du signal audio par le module de contrôle d'inaudibilité et ajoutées au signal audio.

11.2.2 Détection des patterns de synchronisation

Dans cette section, nous nous intéressons à la technique de détection des patterns de synchronisation, qu'il s'agisse de la pattern d'initialisation ou d'acquisition. Ces patterns seront donc notées indifféremment $\mathbf{p}$.

11.2.2.1 Choix du récepteur

Deux types de réception ont jusqu'à maintenant été considérées : celle exploitant un filtre de blanchissant et un dictionnaire de réception et celle basée sur l'égaliseur de Larbi. Pour réaliser la détection des patterns de synchronisation, nous n'utiliserons par la suite que l'égaliseur de Larbi. Deux raisons explicitent ce choix :

- d'une part, l'efficacité de l'égaliseur devant le récepteur par filtre blanchissant en terme de TEB. Cette efficacité laisse présager une meilleure détection des patterns de synchronisation,

- d'autre part, la faible complexité de l'égaliseur par rapport au récepteur par filtre blanchissant. Le récepteur par filtre blanchissant nécessite en effet de calculer un dictionnaire de réception, correspondant aux formes d'onde du dictionnaire d'émission filtrées par une estimation du filtre de mise en forme perceptuelle $\hat{H}(f)$ et par le filtre blanchissant $\hat{G}(f)$. Ce dictionnaire de réception dépendant donc du signal audio analysé, à partir duquel est calculé $H(f)$. Utiliser un tel dictionnaire nécessiterait donc de calculer plusieurs estimations de la pattern de synchronisation à la réception pour chaque fenêtre du signal audio où la pattern de synchronisation est recherchée. Ce procédé est donc très coûteux en temps de calcul. A l'inverse, l'égaliseur de Larbi recherche la pattern de synchronisation sous son

format d'origine (sans avoir besoin de calculer une estimation de cette pattern après filtrage par les filtres de la chaîne). Sa mise en œuvre dans un contexte de synchronisation est donc plus efficace et plus rapide que celle du filtre blanchissant.

Les patterns de synchronisation sont donc détectés en exploitant l'égaliseur de Larbi. Le signal reçu est dans un premier temps filtré par un filtre zero-forcing $\frac{1}{\hat{H}(f)}$ où $\hat{H}(f)$ est le filtre de mise en forme perceptuel calculé à partir du signal audio tatoué détérioré ; le signal résultant $\hat{z}(n)$ est dans un deuxième temps filtré par un filtre de Wiener calculé spécifiquement pour estimer à partir de $\hat{z}(n)$ le signal recherché, c'est à dire ici la pattern de synchronisation. En effet, les coefficients du filtre de Wiener dépendent des propriétés statistiques du signal reçu $\hat{z}(n)$ et du signal émis recherché (précédemment le signal modulé mais ici la pattern de synchronisation). Ces coefficients sont calculés de sorte à minimiser l'erreur quadratique moyenne entre le signal reçu $\hat{z}(n)$ et la pattern ; ils sont donnés sous format vectoriel par :

$$\mathbf{w}(\mathbf{p}) = R_{\hat{\mathbf{z}}}^{-1} r_{\mathbf{p}}, \tag{11.8}$$

où la notation $\mathbf{w}(\mathbf{p})$ est introduite pour mettre en valeur la spécificité des coefficients pour la recherche de $\mathbf{p}$, $R_{\hat{\mathbf{z}}}$ est la matrice d'autocovariance du signal $\hat{z}(n)$ et $r_{\mathbf{p}}$ est la fonction d'autocovariance de la pattern $\mathbf{p}$.

Bien que la recherche des patterns de synchronisation exploitent l'égaliseur de Larbi, la détection des symboles d'information pourra être effectuée aussi bien avec l'égaliseur qu'avec le récepteur par filtre blanchissant.

11.2.2.2 Détection avec prise en compte de la dérive

Supposons pour l'instant que la localisation théorique de la pattern $\mathbf{p}$ est connue.

En l'absence de désynchronisation, le signal reçu $\hat{z}(n)$ après filtrage par $\frac{1}{\hat{H}(f)}$ se décompose sous la forme de la somme entre le signal $v_s(n)$ résultant de la modulation de la pattern de synchronisation et d'un bruit $b(n)$. En se plaçant sur la fenêtre où est insérée la pattern :

$$\hat{\mathbf{z}} = \mathbf{p} + \mathbf{b}. \tag{11.9}$$

Après filtrage par Wiener pour estimer au regard de $\hat{\mathbf{z}}$ la pattern $\mathbf{p}$, une détection de la pattern par une méthode de corrélation par fenêtre glissante peut être effectuée. La corrélation entre le signal reçu (après filtrage de Wiener) $\hat{v}(n)$ et la pattern de synchronisation recherchée $\mathbf{p}$ est calculée au voisinage de la localisation théorique de la pattern. Un pic de la corrélation indique alors l'échantillon de début de la pattern de synchronisation.

Lorsqu'une désynchronisation se produit, par exemple, une opération de "*time stretching*" entraînant une dilation de la fréquence F_e à la fréquence $\hat{F}_e$, le signal reçu n'est plus la pattern bruitée mais une version dilatée de la pattern, notée $\hat{\mathbf{p}}$:

$$\hat{\mathbf{z}} = \hat{\mathbf{p}} + \hat{\mathbf{b}}, \tag{11.10}$$

où $\hat{\mathbf{b}}$ est une version dilatée du bruit $\mathbf{b}$. Le filtre de Wiener et le détecteur par corrélation glissante seront donc d'autant plus efficaces s'ils cherchent à estimer et à récupérer la pattern dilatée $\hat{\mathbf{p}}$ plutôt que la pattern d'origine $\mathbf{p}$. Si la fréquence d'échantillonnage de réception $\hat{F}_e$ est

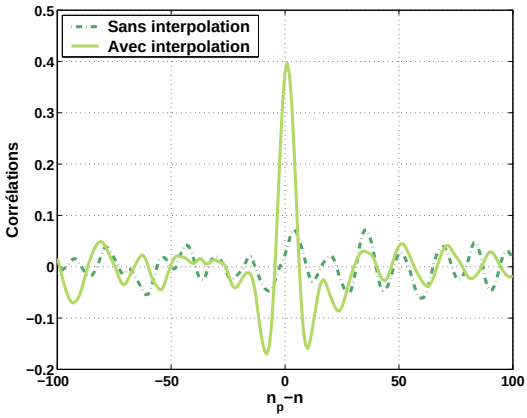


Figure 11.3 – Fonctions de corrélation entre le signal reçu $\hat{v}(n)$ et la pattern de synchronisation au voisinage de la localisation de la pattern lorsque le signal audio tatoué a subi une interpolation par une dérive $d = 2\%$: la courbe en pointillés présente le cas où la pattern $\mathbf{p}$ n'est pas interpolée et le signal $\hat{v}(n)$ est obtenu par un filtre de Wiener spécifiquement adapté à $\mathbf{p}$; la courbe continue présente le cas où la pattern $\hat{\mathbf{p}}$ est interpolée en fonction de la dérive et le signal $\hat{v}(n)$ est extrait du filtrage de Wiener spécifiquement adapté à $\hat{\mathbf{p}}$.

connue, une version interpolée de la pattern de synchronisation $\hat{\mathbf{p}}$ peut être calculée en exploitant la formule d'interpolation établie équation (11.6) :

$$\hat{p}(n) = \sum_{i=-I}^{I} p\left(\left[\frac{F_e}{\hat{F}_e}n\right] + i\right) \sin_c\left(i + \left[\frac{F_e}{\hat{F}_e}n\right] - \frac{F_e}{\hat{F}_e}n)\right), \tag{11.11}$$

où I est la précision sur le calcul de l'interpolation. Les coefficients $\mathbf{w}(\hat{\mathbf{p}})$ du filtre de Wiener peuvent ensuite être spécifiquement calculés pour estimer la pattern dilatée $\hat{\mathbf{p}}$ en appliquant l'équation (11.8) et la méthode de corrélation par fenêtre glissante peut être modifiée pour rechercher dans le signal reçu (après filtrage de Wiener) $\hat{v}$ la pattern interpolée en calculant la corrélation entre $\hat{v}$ et $\hat{\mathbf{p}}$.

Pour justifier de l'efficacité de la prise en compte de la pattern interpolée $\hat{\mathbf{p}}$ dans le mécanisme de synchronisation, nous présentons figure 11.2.2.2 les valeurs de la corrélation entre le signal reçu $\hat{v}(n)$ et la pattern de synchronisation au voisinage de sa localisation lorsque le signal a subi une opération de *time stretching* (simulée par une interpolation) avec une dérive de 2%. Deux cas de figure sont considérés : le premier est celui où la pattern n'est pas interpolée et le second celui où la pattern est interpolée et le filtre de Wiener est spécifiquement calculé pour cette pattern. Cette figure met en évidence que le pic de corrélation, localisant la position de la pattern dans le signal reçu, est nettement plus marqué dans le cas où la pattern est interpolée que dans le cas où elle ne l'est pas. Adapter la détection (le filtre de Wiener et la détection par la corrélation) à l'estimation de la fréquence d'échantillonnage de réception (ou de la dérive) permet donc une meilleure localisation des patterns de synchronisation et donc un meilleur fonctionnement du mécanisme de synchronisation.

11.2.2.3 Sur-échantillonnage

Cette technique de corrélation par fenêtre glissante ne permet de localiser la pattern de synchronisation qu'à l'échantillon près. En effet, étant donnée que le rapport entre la fréquence d'échantillonnage de réception $\hat{F}_e$ et celle de l'émission F_e n'est pas nécessairement entier, l'indice détecté $\hat{n}$ est, d'après l'équation de la dérive temporelle (11.5), l'arrondi de la position effective de la pattern. Pour accroître la précision sur la localisation de la pattern, les mécanismes de synchronisation propose généralement un sur-échantillonnage du signal reçu. Dans le contexte du tatouage (où la fréquence d'échantillonnage de réception est imposée par la perturbation), ce sur-échantillonnage pourrait prendre la forme d'une interpolation pour déterminer la fraction d'échantillon affinant la localisation de la pattern. Des tests [CM04] ont été effectués sur l'apport de cette interpolation ; ils concluent que cette interpolation a très peu d'influence sur les performances du mécanisme de synchronisation. Elle ne sera donc pas considérées par la suite.

11.2.3 Phase d'initialisation

La phase d'initialisation a pour objectif d'estimer la fréquence d'échantillonnage à la réception $\hat{F}_e$ et le retard $\hat{n}_0$ à l'aide des M_i patterns d'initialisation $\mathbf{p}_i$ insérées en amont du signal modulé. Pour des raisons d'implémentation, nous supposerons que les patterns de synchronisation sont recherchées sur une durée N_{ri} limitée en début de signal audio[3].

11.2.3.1 Première phase

Lors de cette phase d'initialisation, $\hat{F}_e$ n'est pas connue. Le signal reçu $\hat{v}(n)$ à partir duquel est estimé la localisation des patterns est donc obtenu après un filtrage de Wiener dont les coefficients sont spécifiquement calculés pour la pattern non interpolée $\mathbf{p}_i$. L'objectif est ensuite de déterminer les M_i pics de corrélation indiquant la localisation des M_i patterns insérées.

La section précédente a mis en évidence la faible corrélation entre le signal reçu $\hat{v}(n)$ et la pattern $\mathbf{p}_i$. Pour accroître les chances de détecter les M_i pics de corrélation, la phase d'initialisation procède donc à une recherche exhaustive des patterns en utilisant plusieurs versions interpolées de celle-ci. P_i versions interpolées de $\mathbf{p}_i$, notées $\{\hat{\mathbf{p}}_i^j\}_{j\in[\![0,P_i-1]\!]}$, sont évaluées pour un ensemble de P_i valeurs de $\hat{F}_e$ couvrant l'intervalle des fréquences de réception tolérées par le système (ici $\hat{F}_e = F_e \pm 2\%$). Pour chaque version interpolée $\hat{\mathbf{p}}_i^j$, la fonction d'intercorrélation $f_j(n)$ entre le signal $\hat{v}(n)$ et $\hat{\mathbf{p}}_i^j$ est calculée. De l'ensemble de ces P_i fonctions doivent être déterminés les M_i pics de corrélation indiquant la position des patterns. Dans cet objectif, une unique fonction d'intercovariance est calculée en sélectionnant échantillon par échantillon le maximum des intercovariances : $f(n) = \max_{j\in[\![0,P_i-1]\!]} f_j(n)$; une illustration de la fonction d'intercovariance obtenue est présentée figure 11.4. De cette fonction sont finalement extraits les M_i pics de corrélation par recherche des maxima. Ces pics donnent accès aux M_i positions des patterns d'initialisation.

3. Un système temps-réel ne devrait pas être limité à cette durée N_r mais devrait rechercher les patterns de synchronisation au fil de la réception du signal audio. La méthode de corrélation par fenêtre glissante comparerait alors la corrélation entre le signal reçu et la pattern de synchronisation à un seuil. Dès que la corrélation dépasserait ce seuil, la localisation de la pattern serait détectée. La présence d'un bruit pourrait entraîner des pics dans la corrélation qui conduirait à de fausses détections de la pattern. Un mécanisme détectant ces fausses détections devrait alors être mis en place. Cette approche nécessite un effort de paramétrisation du système et la mise en place en temps-réel de celui-ci, ce qui n'a pas encore été effectué à ce jour.

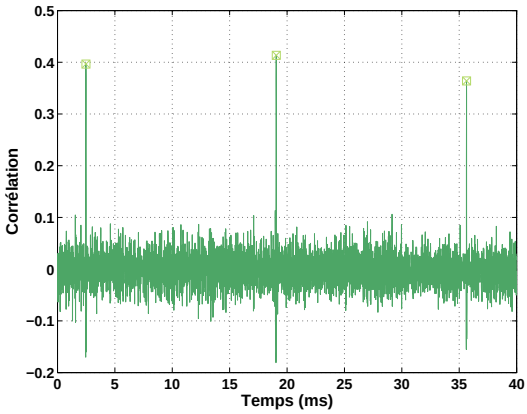


Figure 11.4 – Exemple de fonction d'autocovariance $f(n)$ pour la détection des patterns de synchronisation initiale.

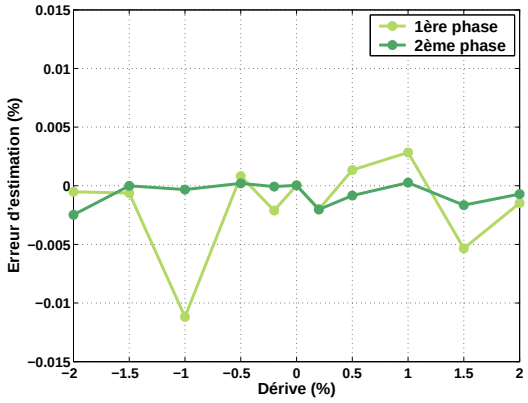


Figure 11.5 – Erreur d'estimation moyenne de la fréquence d'échantillonnage de réception $\hat{F}_e$ en fonction de la dérive et du degré de traitement de la phase d'initialisation.

Connaissant la position des patterns de synchronisation dans le signal reçu, le retard $\hat{n}_0$ introduit par l'opération désynchronisante peut aisément être compensé. De plus, la fréquence d'échantillonnage à la réception peut être évaluée en comparant la distance entre deux patterns de synchronisation dans le signal reçu à leur distance théorique $N_i + N_z$. Cette évaluation donne lieu à une première estimation de $\hat{F}_e$.

11.2.3.2 Seconde phase

Pour affiner cette estimation, un second traitement du signal est effectué. Cette fois, le filtre de Wiener est calculé spécifiquement pour estimer la version interpolée $\hat{\mathbf{p}}_i$ de la pattern de synchronisation initiale à la fréquence $\hat{F}_e$. Le signal reçu $\hat{v}(n)$ est recalculé pour ce nouveau filtre de Wiener et la détection par corrélation glissante est de nouveau effectuée pour $\hat{\mathbf{p}}_i$. M_i positions des patterns en sont déduites, ce qui permet de réestimer le retard $\hat{n}_0$ et la fréquence d'échantillonnage $\hat{F}_e$.

11.2.3.3 Paramétrisation

La performance du mécanisme de synchronisation est fortement dépendante de la paramétrisation du mécanisme de synchronisation (durée des patterns de synchronisation, nombre de fréquences d'échantillonnages P_i considérées, etc...). Cette paramétrisation a été effectuée par une série de tests pratiques en se basant sur l'idée que la détection de la pattern de synchronisation est d'autant meilleure que cette pattern est de durée importante, mais limitée par le compromis débit-robustesse. Nous avons finalement choisi les paramètres suivants : la durée des patterns est de $N_i = 12.N_{HG}$ (où N_{HG} est la taille des fenêtres d'analyse du signal audio pour le calcul des filtres de mise en forme perceptuelle, de Wiener et du filtre blanchissant), elles sont entrecoupées de $N_z = 2N_{HG}$ zéros et elles sont au nombre de $M_i = 3$. La 1ère phase de l'estimation de $\hat{F}_e$ utilise $P_i = 21$ versions interpolées correspondant à un écart entre $\hat{F}_e$ et F_e de $\pm 100.i$ échantillons avec $i \in [\![0, 10]\!]$ (couvrant les dérives de -2 à 2%). Ces paramètres bien

que permettant une synchronisation efficace comme nous allons le voir par la suite, ne sont pas nécessairement les valeurs nominales du mécanisme de synchronisation. Une analyse plus affinée de la paramétrisation devra donc être effectuée dans une étude ultérieure.

11.2.3.4 Performances

Pour juger la performance du mécanisme de synchronisation dans sa phase d'initialisation, la figure 11.5 présente l'erreur relative moyenne sur l'estimation de la fréquence d'échantillonnage de réception en fonction de la dérive ; cette erreur e est le rapport entre l'erreur absolue d'estimation (c'est à dire la différence entre l'estimation moyenne de $\tilde{F}_e$ après tatouage des 20 signaux du corpus et la valeur théorique $\hat{F}_e$ supposée connue) de la fréquence d'échantillonnage de réception et la valeur effective de l'erreur :

$$e = \frac{\tilde{F}_e - \hat{F}_e}{\hat{F}_e}. \tag{11.12}$$

Pour cette simulation, nous avons pris soin de choisir des valeurs de la dérive à estimer, différentes de celles prises en compte par l'algorithme d'estimation [4]. Cette figure met donc en évidence l'efficacité de l'estimation de $\hat{F}_e$ à la réception. La première phase du mécanisme de synchronisation fournit une estimation relativement correcte de $\hat{F}_e$: quelle que soit la dérive l'erreur est inférieure à une dizaine d'échantillons par seconde. La seconde phase réduit cette erreur à environ 1-2 échantillons par seconde.

11.2.4 Phase d'acquisition

La phase d'acquisition consiste à détecter les patterns de synchronisation $\mathbf{p}_a$ insérées régulièrement dans le signal de tatouage pour permettre de synchroniser le récepteur et affiner l'estimation de la fréquence d'échantillonnage de réception $\hat{F}_e$. Ces patterns servent en effet de points de référence marquant le début de chaque message.

11.2.4.1 Détection des patterns d'acquisition

Considérons la détection de la pattern $\mathbf{p}_a$ en amont du j-ème message. S'il s'agit du premier message faisant suite au pattern de synchronisation initiale $\mathbf{p}_i$, sa localisation théorique $\hat{n}_0$ est aisément déduite : en effet, l'étape d'initialisation a permis d'estimer la position $\hat{n}_i$ de la première pattern $\mathbf{p}_i$ et la fréquence d'échantillonnage $\hat{F}_e$. Connaissant la distance théorique $M_i(N_i + N_z)$ séparant la première pattern d'initialisation et la première pattern de synchronisation, nous en déduisons :

$$\hat{n}_0 = \hat{n}_i + \left[M_i(N_i + N_z)\frac{\hat{F}_e}{F_e} \right]. \tag{11.13}$$

S'il s'agit du j-ème message, on connaît la position $\hat{n}_{j-1}$ de la pattern d'acquisition du $(j-1)$-ème message ; sachant la distance N_m séparant ces deux patterns et l'estimation de $\hat{F}_e$, nous obtenons :

$$\hat{n}_j = \hat{n}_{j-1} + \left[N_m \frac{\hat{F}_e}{F_e} \right]. \tag{11.14}$$

4. En effet, l'algorithme d'estimation prend en compte les fréquences d'échantillonnage $\hat{F}_e = F_e \pm 100.i$, tandis que les fréquences d'échantillonnage recherchées prennent pour valeurs possibles : 43218, 43438, 5, 43659, 43879, 5, 44100, 44320, 5, 44541, 44761, 5 et 44982.

Connaissant la position théorique $\hat{n}_j$ de la pattern d'acquisition du j-ème message, la phase d'acquisition consiste à rechercher la position exacte de cette pattern sur un voisinage de sa location théorique de durée $N_r a$. Le signal reçu $\hat{z}(n)$ après filtrage par $\frac{1}{\hat{H}(f)}$ est traité par un filtre de Wiener dont les coefficients sont spécifiquement calculés pour estimer la version interpolée, notée $\hat{\mathbf{p}}_a$, de $\mathbf{p}_a$ à la fréquence $\hat{F}_e$. Le signal résultant $\hat{v}(n)$ est ensuite examiné par la méthode de corrélation par fenêtre glissante : la fonction d'intercovariance entre $\hat{v}(n)$ et la pattern interpolée $\hat{\mathbf{p}}_a$ est calculée sur le voisinage de $\hat{n}_j$. La présence d'un pic dans la fonction d'intercovariance indique la position exacte de la pattern $\hat{n}_j$.

11.2.4.2 Actualisation de $\hat{F}_e$

Détecter la pattern d'acquisition en début du j-ème message permet également d'affiner l'estimation de $\hat{F}_e$. Deux méthodes d'actualisation ont été considérées :

- **Méthode n°1** : $\hat{F}_e$ peut être actualisée en comparant la distance mesurée entre les deux dernières patterns d'acquisition détectées à la distance théorique N_m qui les sépare :

$$\hat{F}_e = \frac{\hat{n}_j - \hat{n}_{j-1}}{N_m} F_e. \tag{11.15}$$

- **Méthode n°2** : $\hat{F}_e$ est réévaluée en comparant la distance qui sépare la pattern de synchronisation du j-ème message et la la pattern de synchronisation du 1er message à leur distance théorique :

$$\hat{F}_e = \frac{\hat{n}_j - \hat{n}_0}{jN_m} F_e. \tag{11.16}$$

11.2.4.3 Paramétrisation

Là encore, l'efficacité du mécanisme de synchronisation est fortement liée au choix des paramètres (durée des patterns d'acquisition, taille du message, etc...). Les choix que nous avons effectués proviennent de tests sur des signaux réels, et ne prétendent pas être les valeurs nominales pour la synchronisation.

La taille des messages N_m dépend de la taille des patterns d'acquisition N_a, de la durée du temps symbole N_s et du nombre de symboles L_m par message. Elle fixe en quelque sorte la fréquence de répétition des patterns d'acquisition et impose le débit de fonctionnement interne du système (qui diffère du débit utile). Les patterns d'acquisition sont choisies de durée $N_a = 15N_{HG}$ [5]. Cette durée est du même ordre de grandeur que la durée des patterns d'initialisation ; elle permet donc une estimation relativement fiable de la localisation des patterns. N_m ne doit être choisi ni trop court pour éviter une diminution trop importante du débit de transmission effectif ni trop long pour garantir une synchronisation régulière du récepteur sur les patterns. Nous avons choisi une taille de message N_m d'environ F_e échantillons (1 seconde de signal) : N_m peut en effet légèrement varié autour de F_e en fonction de la contrainte de débit de transmission utile $R = \frac{KF_e}{N_s}$. Cette contrainte impose le temps symbole N_s ; le nombre de symboles émis dans

5. Ce choix relève, comme dans le cas des patterns de synchronisation initiale, de l'idée que la durée des patterns doit être choisie suffisamment longue pour en permettre la détection aisée mais doit être limitée par le compromis débit-robustesse. Cette durée est également choisie être multiple de la taille des fenêtres N_{HG} du calcul des filtres pour faciliter l'implémentation du mécanisme de synchronisation.

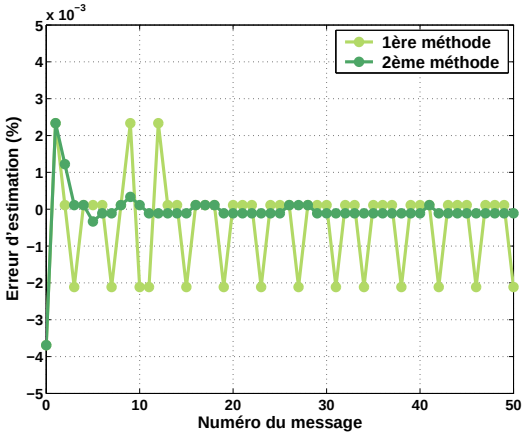


Figure 11.6 – Évolution de l'erreur d'estimation sur la fréquence d'échantillonnage à la réception $\hat{F}_e$ pour deux méthodes de suivi lorsque la dérive est de 2%.

un message est ensuite déduit $L_m = \left\lceil \frac{N_m - N_a}{N_s} \right\rceil$ où $\lceil x \rceil$ désigne la partie supérieure de x, avec pour l'instant $N_m = F_e$; et la valeur effective de $N_m = N_a + L_m N_s$ est calculée.

11.2.4.4 Performances

Les performances de la phase d'acquisition sont liées à sa capacité à estimer $\hat{F}_e$ au cours du temps. En effet, si les patterns sont mal localisées, $\hat{F}_e$ sera erronée. La figure 11.6 présente l'évolution de l'erreur d'estimation sur la fréquence $\hat{F}_e$ obtenue au fur et à mesure du traitement du signal reçu. Ce signal résulte du tatouage d'un des signaux de la base test à un débit de 100 bit/s. Les résultats seront identiques à d'autres débits, puisque le fonctionnement du mécanisme d'acquisition de la fréquence $\hat{F}_e$ est indépendant du débit.

Cette figure met en évidence l'efficacité du mécanisme de synchronisation. Les deux méthodes d'actualisation permettent de converger vers une erreur d'estimation sur $\hat{F}_e$ quasi nulle. La méthode n°2 semble plus efficace que la première. En effet, la première méthode est plus sensible à des imprécisions sur la localisation des patterns que la seconde. Par la suite, nous ne conserverons donc que la méthode d'actualisation n°2.

11.2.5 Phase de poursuite

11.2.5.1 Mécanisme de synchronisation locale

La pattern de synchronisation en amont du j-ème message ayant été localisée à l'instant $\hat{n}_j$ et l'estimation de $\hat{F}_e$ ayant été faite (ou actualisée), il est alors possible d'estimer la localisation théorique de chaque symbole d'information faisant suite à la pattern et de procéder à leur détection. Cette détection peut être effectuée aussi bien pour un récepteur basé sur l'égaliseur de Larbi que sur le récepteur par filtre blanchissant. En effet, la détection des patterns de synchronisation étant effectuée indépendamment de la détection de l'information cachée, il est tout à fait possible d'utiliser un récepteur différent pour la recherche des patterns et pour la détection.

Le i-ème symbole du j-ème message débute à l'instant $\hat{n}_j^i$ avec :

$$\hat{n}_j^i = \hat{n}_j + \frac{\hat{F}_e}{F_e}(N_a + iN_s). \tag{11.17}$$

La localisation du i-ème symbole reste malheureusement imprécise : d'une part, l'instant marquant le début de la pattern de synchronisation ne peut être déterminé à l'échantillon près à moins d'interpoler sur une fraction d'échantillon. D'autre part, si la dérive d est mal estimée, la position des symboles est également mal évaluée. Leurs positions effectives par rapport à la position théorique sont d'ailleurs d'autant plus éloignées que le symbole détecté est loin de la patten de synchronisation ; la taille des messages N_m ne devra donc pas être trop importantes.

Il est donc nécessaire d'introduire un mécanisme de synchronisation locale qui permet de synchroniser le récepteur sur chaque temps symbole. Ce mécanisme est basé sur une technique "*early-late gate*" [Pro01] qui affine la localisation du symbole et procède à sa détection simultanément. Cette technique exploite de nouveau un corrélateur par fenêtre glissante sur un voisinage N_d de la position théorique de chaque symbole. Pour chaque symbole, la valeur de la corrélation entre le signal reçu et chaque vecteur du dictionnaire de réception est calculée sur le voisinage N_d de $\hat{n}_j^i$. La valeur maximale de la corrélation indique à la fois l'instant de début du symbole et la valeur du symbole (puisque la détection dans le cas d'un système de tatouage exploitant des modulations M-aire est basé sur la recherche du maximum de corrélation).

11.2.5.2 Dictionnaire de réception

Comme dans le cas de la détection des patterns de synchronisation, la détection est d'autant meilleure que le dictionnaire de réception ($\mathcal{D}$ dans le cas d'un récepteur par égaliseur de Wiener et $\hat{\mathcal{D}}$ dans le cas du filtre blanchissant) reflète les distorsions introduites par la désynchronisation. Pour être efficace, le dictionnaire $\bar{\mathcal{D}}$ utilisé par le corrélateur contient l'ensemble des vecteurs du dictionnaire de réception interpolés à la fréquence estimée $\hat{F}_e$.

11.2.5.3 Spécificités de la réception pour l'égaliseur de Larbi

Outre le fait d'utiliser pour la démodulation par corrélation un dictionnaire "interpolé", la réception pour l'égaliseur de Larbi gagne à utiliser un filtre de Wiener spécifiquement adapté au signal à estimer. Ici, il s'agit des formes d'onde du dictionnaire d'émission, ayant subie une interpolation à une fréquence $\hat{F}_e$. Le calcul des coefficients du filtre de Wiener pour la détection des patterns de synchronisation est donc effectué à partir d'un signal modulé construit à l'aide du dictionnaire interpolé.

11.2.5.4 Paramétrisation

La paramétrisation de ce mécanisme de synchronisation locale tient au choix du voisinage N_d utilisé pour la technique de corrélation par fenêtre glissante mais également au choix de la taille de message N_m. Si l'on fait l'hypothèse que l'erreur d'estimation de $\hat{F}_e$ est relativement faible, nous pouvons autorisé un voisinage N_d faible (de 1 échantillon) mais choisir une longueur de message relativement faible pour se prémunir contre l'accroissement de l'erreur sur la localisation du symbole avec sa distance à la pattern de synchronisation. Cette constatation justifie une nouvelle fois le choix d'un message d'environ 1 seconde ($N_m \approx F_e$), qui constitue le bon compromis entre faible longueur du message et accroissement du débit utile.

11.3 Performances

Nous proposons dans cette section d'évaluer l'efficacité du mécanisme de synchronisation sur la robustesse du système de tatouage aux attaques désynchronisantes. Pour ce faire nous avons donc considéré les deux stratégies d'insertion les plus performantes proposées dans ce manuscrit : il s'agit de la stratégie d'insertion informée avec dictionnaire structuré et choix d'un tatouage de robustesse maximales aux perturbations décrite section 9.4 pour les modulations M-aire. Les deux variantes de cette stratégie seront considérées : celle avec le récepteur par filtre blanchissant et celle basée sur l'égaliseur de Larbi.

Pour évaluer la robustesse du système, nous avons soumis les signaux audio tatoués à différents débits (83 bit/s, 165 bit/s et 245 bit/s) à une opération de *time stretching* de dérive variable (supposée être au maximum de $\pm 2\%$), simulée à l'aide de la formule d'interpolation donnée équation (11.6). Un retard de 1000 échantillons a également été introduit sur les fichiers audio reçus. La figure 11.7 présente les résultats obtenus : les courbes (a) et (c) rappellent les performances des deux systèmes dans le cas d'un canal sans perturbation. Les courbes (b) et (d) proposent les variations des TEBs en fonction de la dérive et pour les différentes valeurs de débit considérées lorsque le mécanisme de synchronisation est activé pour les deux systèmes informés : celui basé sur le récepteur par filtre blanchissant dans le cas de la figure (b) et celui basé sur l'égaliseur de Larbi dans le cas de la figure (d).

Au regard de ces courbes, nous pouvons constaté une augmentation des TEBs lorsque le mécanisme de synchronisation est mis en œuvre. Dans le cas de l'égaliseur de Larbi, le TEB pour un débit de 165 bit/s passe de 8.10^{-3} (pour le système sans désynchronisation) à $1,7.10^{-2}$ lorsque la dérive est nulle, soit une augmentation par 2 des TEBs. Par contre, la robustesse du système de tatouage apparaît indépendante de la dérive : les TEBs sont quasiment constants lorsque le débit est supérieur à 165 bit/s. Des variations plus importantes sont observées pour un débit de 85 bit/s. Néanmoins, les TEBs atteints pour ce débit de transmission sont inférieurs à 10^{-3}, c'est à dire à la fiabilité des mesures proposées dans ce manuscrit. Les mesures présentées dans cette configuration de débit ne sont donc pas significatives. Dans le cas du récepteur par filtre blanchissant, l'augmentation du TEB liée à l'activation du mécanisme de synchronisation est moins importante que dans le cas de l'égaliseur. A 165 bit/s, le TEB passe de $1,9.10^{-2}$ (pour le système sans désynchronisation) à 2.10^{-2}. Par contre, les variations du TEB en fonction de la dérive sont nettement plus marquées. Dans le pire des cas (ici une dérive de -2%), les TEBs sont multipliés par 2 par rapport au cas sans dérive. Les deux récepteurs présentent donc un comportement différent face à la perturbation désynchronisante ; le système informé basé sur l'égaliseur de Larbi reste pourtant plus efficace que celui basé sur le récepteur par filtre blanchissant, comme cela était le cas sans activation du mécanisme de synchronisation.

Le mécanisme de synchronisation est donc relativement efficace dans ce sens qu'il limite la dégradation des TEBs. Dans le pire des cas, les TEBs sont multipliés par 2 par rapport aux TEBs du système sans mis en œuvre du mécanisme de synchronisation. On pourra noter que cette augmentation par 2 est du même ordre de grandeur que celle induit par la perturbation non désynchronisante la plus dégradante (l'ajout de bruit), que nous avons exposé section 9.4.2.

Nous nous sommes également intéressés au coût, en terme de temps de calcul, du mécanisme de synchronisation. Ce coût est mesuré comme le rapport entre le temps nécessaire au traitement de la réception et la durée du signal audio traité. La table 11.1 présente les résultats obtenus par le système lorsque le mécanisme de synchronisation est activé et ne l'est pas[6]. L'activation

6. Les mesures présentées ne prennent pas en compte la phase d'initialisation de $\hat{F}_e$ puisque celle-ci n'est

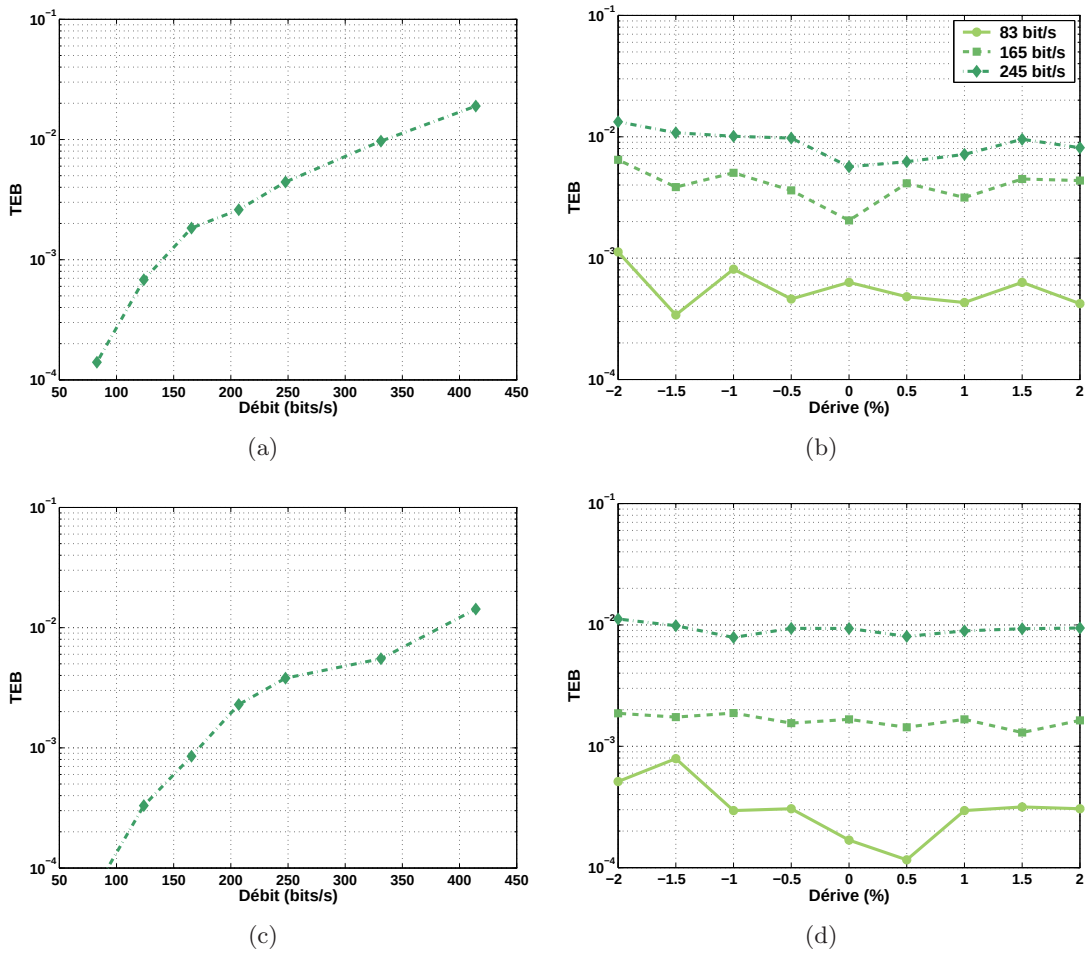


Figure 11.7 – Mise en évidence des performances du mécanisme de synchronisation pour deux récepteurs : (a) TEBs du système avec réception par filtre blanchissant dans le cas d'un canal sans perturbations, (b) TEBs du système avec réception par filtre blanchissant en fonction de la dérive et du débit, (c) TEBs du système avec l'égaliseur de Larbi dans le cas d'un canal sans perturbations et (d) TEB du système avec l'égaliseur de Larbi en fonction de la dérive et du débit.

du mécanisme de synchronisation, comme nous pouvions nous y attendre, augmente le coût de la réception : cette augmentation est relativement faible dans le cas du récepteur par filtre blanchissant (le coût y est multiplié par environ 2). Elle est ici essentiellement liée à l'augmentation du nombre de calculs de corrélations. Par contre, elle est nettement plus importante dans le cas de l'égaliseur : en effet, le mécanisme de synchronisation nécessite cette fois, en plus des calculs de corrélation supplémentaires, l'estimation de la matrice d'autocovariance du signal modulé sur chaque message. Les performances du mécanisme de synchronisation, puisqu'étant liées au nombre de calculs de corrélations, sont naturellement dépendantes du nombre de vecteurs du dictionnaire. Rappelons que dans cette configuration de mesure, ce nombre est relativement

effectuée qu'une seule fois ; elles concernent la phase d'acquisition et de poursuite qui constitue le régime courant du système de tatouage.

Récepteur	Filtre blanchissant	Égaliseur de Larbi
Sans mise en œuvre de la synchronisation	409%	40%
Avec mise en œuvre de la synchronisation	739%	283%

Table 11.1 – Coût en temps de calcul par rapport au temps-réel des deux récepteurs lorsque le mécanisme de synchronisation est ou non activé.

grand ($M = 64$ vecteurs). Si l'on réduit le dictionnaire à deux vecteurs, le coût de calcul du récepteur (avec le filtre blanchissant ou avec l'égaliseur de Larbi) est décru sous la limite des 100%. La faisabilité du système de tatouage disposant du mécanisme de synchronisation peut donc être atteinte, à condition de choisir un nombre de vecteurs du dictionnaire relativement faible.

Conclusion

Dans cette section, nous avons proposé un mécanisme de synchronisation adapté à toutes opérations de désynchronisation, à condition qu'elles puissent être modélisées par un rééchantillonnage du signal audio tatoué à une fréquence constante. Il est basé sur l'ajout de patterns de synchronisation, permettant d'estimer la dérive d'échantillonnage, même forte, pour adapter les paramètres du récepteur. Des résultats expérimentaux sur des signaux réels ont montré que la fiabilité de transmission obtenue pour un canal sans perturbation est faiblement dégradée si le canal devient une désynchronisation à forte dérive, attestant ainsi de la performance du mécanisme proposé. Une augmentation par 2 des TEBs est à prévoir pour une dérive allant jusqu'à $\pm 2\%$. Pour obtenir une transmission de fiabilité 10^{-3} même face à une opération de "*time stretching*", le débit de transmission doit notamment être choisi aux alentours de 125 bit/s (contre 175 bit/s pour un canal non désynchronisant). Le système peut également être implémenté en temps-réel à condition de choisir des paramètres d'encodage de la séquence binaire relativement faibles.

Conclusions et perspectives

Ce manuscrit traite du tatouage des signaux audio numériques dans le contexte particulier de la transmission de données cachées. Cette classe applicative permet notamment d'augmenter le contenu d'un signal audio ou d'utiliser ce signal comme un canal de transmission lorsque le format de diffusion (mp3, analogique, etc...) du signal n'est pas garanti.

La finalité première de cette thèse s'attache à la conception d'un tel système de tatouage. Cette conception est régie par quatre contraintes, qui établissent la notion de performance du système : l'insertion du tatouage doit être transparente, la fiabilité de transmission la meilleure possible pour un débit de transmission le plus élevé possible et ce en présence de toutes les opérations licites pouvant dégrader le signal audio et pour finir le coût de calcul suffisamment faible pour permettre d'envisager l'application temps-réel. L'étude de l'existant a permis un état des systèmes proposés dans la littérature et de leur performance. Elle soulève l'importance du choix du domaine d'insertion et de détection, tant sur le plan de la transparence du tatouage que de la fiabilité de la transmission, mais révèle surtout la nécessité de prendre en compte la connaissance du signal audio durant la phase d'émission : la construction du tatouage relève alors d'une stratégie dite informée, dont le but est de choisir le signal de tatouage garant d'une transmission inaudible et de robustesse maximale aux bruits de perturbations. La difficulté tient alors à la modélisation des contraintes d'inaudibilité du tatouage et robustesse. Plusieurs stratégies proposées dans la littérature montrent l'efficacité d'une telle approche sur les performances du système, lorsque l'accent est mis sur la robustesse au détriment de la contrainte d'inaudibilité. L'objectif de recherche de cette thèse et la contribution majeure de ce manuscrit se sont à l'inverse portés sur la conception de stratégies d'insertion informées dans lesquelles la robustesse est maximisée sous contrainte de l'inaudibilité.

Fort des travaux menés lors de collaborations entre l'ENST, l'Université Paris V et l'ENIT [7], nous nous sommes attachés dans un premier temps à la conception, la formalisation et la justification d'un système de tatouage non informé. L'insertion s'effectue de manière additive dans le domaine temporel. L'inaudibilité est atteinte via l'utilisation d'un modèle psychoacoustique (permettant la mise en forme spectrale du tatouage) et d'un facteur de gain paramétrable. Nous avons proposé un module de contrôle adaptatif du gain au principe innovant. Ce module mesure la transparence du tatouage grâce à l'algorithme PEAQ d'évaluation objective de la différence perceptuelle entre deux signaux (ici le signal audio original et sa version tatouée) et choisi le gain spécifiquement adapté au signal audio pour obtenir un niveau de transparence donné, fixé par le concepteur du système. Le récepteur blanchit le signal reçu avant de procéder à une démodulation par corrélation, adaptée du récepteur théorique optimal. Cette étude a permis de conclure à l'utilité du module de contrôle sur la transparence du tatouage mesurée par l'algorithme PEAQ et d'en évaluer l'efficacité et les défauts. Elle a également établi une mesure de

7. École Nationale d'Ingénieurs de Tunis

référence de la fiabilité de transmission en fonction du débit, tout à fait comparable à celles des systèmes non informés dans l'état de l'Art. Elle a montré la robustesse du système de tatouage aux perturbations non désynchronisantes et à la faisabilité du tatouage en temps-réel.

Nous avons dans un deuxième temps construit un système informé à partir de ce système de référence. Une copie locale du récepteur à l'émetteur, conférant au système une structure en boucle fermée, a été introduite pour prendre en compte la connaissance du signal audio à l'émetteur. Elle permet de formaliser les conditions à satisfaire par le tatouage pour garantir une transmission inaudible, correcte et robuste de l'information face à un bruit de perturbation. Le choix du signal de tatouage a alors été rattaché à plusieurs problèmes d'optimisation recherchant le tatouage de robustesse maximale aux perturbations sous contrainte de l'inaudibilité. Ces problèmes d'optimisation dépendent de la modélisation du bruit (sous la forme d'un paramètre à maximiser ou d'une probabilité d'erreur de transmission face à ce bruit à atteindre), de la modulation dont est issu le tatouage et du récepteur choisi pour sa détection. Ces problèmes d'optimisation ont été résolus en utilisant différents algorithmes. Leurs performances ont d'abord été étudiées sur un modèle théorique de la chaîne de tatouage puis sur le système de tatouage en boucle fermée. Cette étude a permis de conclure à l'efficacité des stratégies d'insertion informées proposées. La transparence du tatouage est très similaire à celle du système de référence. La fiabilité de transmission a été très nettement améliorée : un TEB, estimateur de la probabilité d'erreur de transmission, de 10^{-3} peut désormais être obtenu à un débit de transmission de 170 bit/s alors qu'il nécessitait une transmission à 80 bit/s dans le cas du système de référence. La robustesse du système a été correctement maintenue par rapport au système de référence : une augmentation par 3 des TEBs doit être prévue pour la majorité des perturbations et par $7,5$ dans le cas de la perturbation la plus dégradante, à savoir le filtrage passe-haut. Le coût en temps de calcul du système a par contre été fortement accru notamment au niveau de l'émission qui nécessite une puissance de calcul relativement importante. Une application temps-réel ne peut être pour l'instant envisagée qu'au niveau de la réception.

La dernière contribution de ce manuscrit s'est attachée à la robustesse du système aux perturbations désynchronisantes de type dilation ou contraction de l'échelle des temps dont le taux de dilation/contraction (la dérive) est élevé mais constant (de l'ordre de $\pm 2\%$). Nous avons proposé un mécanisme de synchronisation adapté et peu coûteux en temps de calcul. Ce mécanisme se base sur l'ajout de patterns de synchronisation permettant d'estimer la dérive, même forte, et d'adapter les paramètres du récepteur en conséquence. Une mise en œuvre expérimentale sur des signaux réels a permis d'évaluer les performances de la synchronisation : une augmentation des TEBs de l'ordre de 2 est à prévoir lorsque le mécanisme de synchronisation est activé. Par contre, le TEB obtenu est quasiment indépendant de la dérive. Pour conserver une transmission de TEB 10^{-3} il faudra donc se placer à un débit aux alentours de 125 bit/s.

Cette étude met en lumière de nombreuses améliorations possibles des systèmes de tatouage, tant sur l'inaudibilité, sur la fiabilité de transmission que sur la robustesse aux perturbations.

La mise en œuvre du module de contrôle adaptatif de l'inaudibilité basé sur l'algorithme PEAQ a montré la faisabilité d'imposer à l'émetteur une transparence de tatouage donnée. Elle soulève pourtant l'adéquation de l'algorithme PEAQ construit dans un contexte de compression au contexte du tatouage. Une étude spécifique des dégradations introduites par le tatouage, combinant mesures objectives et mesures subjectives, pourrait être effectuée pour proposer un algorithme spécifiquement dédié au tatouage.

La fiabilité de transmission, pour un système additif, dépend fortement de la projection du signal audio dans l'espace de détection : plus cette projection s'approche d'un bruit blanc,

de faible puissance et décorrélé du tatouage, meilleure est la détection. Or cet espace est issu d'un dictionnaire à partir duquel est construit le tatouage. Il pourrait donc être envisagé de choisir un dictionnaire "informé", adapté au signal audio, choisi pour minimiser la puissance de la projection audio. Des outils de décorrélation seraient utilisés à l'émetteur sans modifier la structure du récepteur, hormis pour réestimer le dictionnaire utilisé.

L'étude des stratégies d'insertion informées pourrait être poursuivie vers la conception de stratégies à débit variable atteignant cette fois une transparence de tatouage fixée et une robustesse aux perturbations donnée, toutes deux choisies par le concepteur du système. Les stratégies proposées jusqu'à maintenant imposent un débit de transmission fixe qui oblige d'établir un compromis entre l'inaudibilité et la robustesse : on sait minimiser l'inaudibilité sous contrainte de robustesse ou à l'inverse maximiser la robustesse sous contrainte d'inaudibilité. Des systèmes de transmission à débit variable pourraient être envisagés. Le tatouage serait alors choisi sous contrainte d'inaudibilité et transmis sur une durée de temps symbole adéquate permettant d'atteindre un niveau de robustesse donné. La difficulté de cette stratégie réside essentiellement dans la capacité du récepteur à estimer la durée de transmission de chaque information.

Une autre perspective serait de combiner les techniques de codage informé (telles que la STDM [Che00]), connues pour permettre un débit de transmission élevé, et les apports de notre système de tatouage en matière d'inaudibilité. La STDM propose une quantification de la projection du signal audio sur un dictionnaire d'émission en fonction de l'information à émettre ; le tatouage prend alors la forme d'un bruit blanc qui peut introduire des distorsions auditives locales. Nous pourrions envisager un système hybride qui exploite un modèle psychoacoustique et le module de contrôle d'inaudibilité de notre système pour réaliser la mise en forme spectrale du tatouage STDM et garantir ainsi son inaudibilité. Plus généralement, cette étude serait l'occasion de poursuivre la comparaison des avantages et des inconvénients des techniques par étalement de spectre additive et des techniques substitutives de codage informé, qui fait encore débat aujourd'hui.

Bibliographie

[56299] Recommandation UIT-R B.S. 562 : Evaluation subjective de la qualité du son, 1999.

[Arn02] Mickael Arnold : Subjective and objective quality evaluation of watermarked audio tracks. Dans *2d International Conference on Web Delivering of Music (Webdelmusic)*, pages 161–167, Darmstadt, Germany, december 2002.

[AS02] Michael Arnold et Kai Schilz : Quality evaluation of watermarked audio tracks. Dans *SPIE Electronic Imaging : Security and Watermarking of Multimedia Content IV*, volume 4675, pages 91–101, San Jose, CA, USA, january 2002.

[AWS03] Michael Arnold, Stephen Wolthusen et Martin Schmucker : *Techniques and applications of digital watermarking and content protection.* Artech House Publishers, july 2003.

[Bar02] Cléo Baras : Etude de la mise en forme de l'information binaire dans un système de tatouage audio. Mémoire de D.E.A., Institut National Polytechnique de Grenoble (INPG), Grenoble, France, août 2002.

[Bar03a] Mauro Barni : What is the future for watermarking ? (part i). *IEEE Signal Processing Magazine*, 20(5):55–60, september 2003.

[Bar03b] Mauro Barni : What is the future for watermarking ? (part ii). *IEEE Signal Processing Magazine*, 20(6):53–59, november 2003.

[Bas00] Patrick Bas : *Méthodes de tatouage d'images fondées sur le contenu.* Thèse de doctorat, Institut National Polytechnique de Grenoble (INPG), Grenoble, France, 2000.

[Bat05] Haïkel Battikh : Modèles psychoacoustiques dans une chaîne de tatouage audio. Rapport technique, (Fin d'études ingénieur), Ecole Polytechnique de Tunis, mai 2005.

[Bau02] Séverine Baudry : *Stratégie de codage de canal pour le tatouage de vidéo.* Thèse de doctorat, École Nationale Supérieure des Télécommunications (ENST), Paris, France, janvier 2002.

[BBC+04] Patrick Bas, Séverine Baudry, François Cayre, Jean-Marc Chassery, Gouenou Coatrieux, Franck Davoine, Damien Delanny, Caroline Fontaine, Teddy Furron, Benoît Macq, Henri Maître, Nicolas Moreau, Philippe Nguyen, Séphane Pateux, Fabien Petitcolas et Francis Schmitt : *Tatouage de documents audiovisuels numériques.* Traitement du signal et de l'image : Information - Commande - Communication. Lavoisier - Hermes Science, 2004.

[BMGL96] Walter Bender, Norishige Morimoto, Daniel Gruhl et Anthony Lu : Techniques for data hiding. *IBM System Journal*, 25:313–335, 1996.

[BMJ95] P. Bourcet, D. Masse et B. Jahan : Système de diffusion de données, 1995. Brevet d'Invention 95 06727, Télédiffusion de France.

[BPG02] Félix Balado et Fernando Pérez-González : Provably or probably robust data hiding ? Dans *IEEE International Conference on Multimedia and Expo (ICME)*, volume 2, pages 481–484, Lausanne, Switzerland, august 2002.

[BPN01] Paraskevi Bassia, Ioannis Pitas et Nikos Nikolaidis : Robust audio watermarking in the time domain. *IEEE Transactions on Multimedia*, 3(2):232–241, june 2001.

[Bra87] Karlheinz Brandenburg : Evaluation of quality for audio encoding at low bit rates. Dans *82nd Convention of Audio Engineering Society (AES)*, London, England, march 1987. Preprint 2433.

[Bro97] Jean-Marc Brossier : *Signal et communication numérique : égalisation et synchronisation.* Hermès, Paris, France, 1997.

[BS94] John G. Beerends et Jan A. Stemerdink : A perceptual speech quality measure based on a psychoacoustic sound representation. *Journal of Audio Engineering Society*, 42:115–123, march 1994.

[BTH96] Laurence Boney, Ahmed H. Tewfik et Khaled N. Hamdy : Digital watermarks for audio signal. Dans *IEEE International Conference on Multimedia Computing and Systems (ICMCS)*, pages 473–490, Hiroshima, Japan, june 1996.

[BvdVL04] Sylvain Beauget, Michel van der Veen et Aweke Lemma : Informed detection of audio watermark for resolving playback speed modifications. Dans *ACM Workshop on Multimedia and Security*, pages 117–123, Magdeburg, Germany, september 2004.

[Cha96] Maurice Charbit : *Eléments de théorie du signal : Aspects aléatoires.* Editions Ellipses, 1996.

[Cha03] Maurice Charbit : *Systèmes de communications et théorie de l'information.* Hermès Science Publications, 2003.

[Che00] Brian Chen : *Design and Analysis of Digital Watermarking, Information Embedding, and Data Hiding Systems.* Thèse de doctorat, MIT, Cambridge, MA, USA, june 2000.

[Cia98] P. G. Ciarlet : *Introduction à l'analyse numérique matricielle et à l'optimisation.* Dunod, Collection Mathématiques appliquées pour la maîtrise, 1998.

[CKTLS97] Ingemar J. Cox, Joe Kilian, F Thomson Leighton et Talal Shamoon : Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1687, december 1997.

[CM04] Virginie Coder et Sylvie Mouton : Méthode de synchronisation dans un système de tatouage audio. Rapport technique, (Fin d'études ingénieur), Ecole Nationale Supérieure d'Electronique et de RadioElectricité (ENSERG), INP Grenoble, juin 2004.

[CMB02] Ingemar Cox, Matthew Miller et Jeffrey Bloom : *Digital watermarking.* Morgan Kaufmann Publishers, San Francisco, USA, 2002.

[CMM99] Ingemar J. Cox, Matthew L. Miller et Andrew L. McKellips : Watermarking as communications with side information. *Proceedings of the IEEE (USA)*, 87(7):1127–1141, july 1999.

[Col94] Catherine Colomes : *Etude d'un modèle d'audition et d'une mesure objective de la qualité d'un signal sonore dans le contexte de codage à réduction de débit.* Thèse de doctorat, Université de Rennes 1, Rennes, France, septembre 1994.

[Cos83] Max Costa : Writing on dirty paper. *IEEE Transactions on Information Theory*, 29:439–441, may 1983.

[CPGR00] Jim Chou, Sandeep Pradhan, Laurent Ghaoui et Kannan Ramchandran : A robust optimization solution to the data hiding problem using distributed source principles. Dans *SPIE conference on Electronic Imaging : Image and Video Communications and Processing*, volume 3974, pages 270–279, San-Jose, California, USA, january 2000.

[CS00] Brian Chen et Carl-Eric W. Sundberg : Digital audio broadcasting in the fm band by means of contiguous band insertion and precanceling techniques. *IEEE Transactions on Communications*, 48(10):1634–1637, october 2000.

[CS01] Qiang Cheng et Jeffrey Sorensen : Spread spectrum signaling for speech watermarking. Dans *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1337–1340, Salt Lake City, UT, USA, may 2001.

[CS03] Nedeljko Cvejic et Tapio Seppänen : Increasing robustness of an improved spread spectrum audio watermarking method using attack characterization. Dans *IEEE International Conference on Information Technology : Coding and Computing*, pages 533–537, Las Vegas, NV, USA, april 2003.

[CTS03] Nedeljko Cvejic, Djordje Tujkovic et Tapio Seppänen : Increasing robustness of an audio watermark using turbo codes. Dans *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1217–1220, Baltimore, MD, USA, july 2003.

[CW01] Brian Chen et George Wornell : Quantization index modulation : a class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, 47:1423–1443, may 2001.

[CWL01] Scott Craver, Min Wu et Bede Liu : What can we reasonably expect from watermarks ? Dans *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 223–226, Mohonk, NY, USA, october 2001.

[CWLF01] Scott Craver, Min Wu, Bede Liu et E. Felten : Analysis of attacks on SDMI audio watermarks. Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1369–1372, Salt Lake City, Utah, USA, may 2001.

[DKL98] Geert Depovere, Ton Kalker et Jean-Paul Linnartz : Improved watermark detection using filtering before correlation. Dans *IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 430–434, Chicago, IL, USA, october 1998.

[Doë01] Gwenaël Doërr : Robust and high capacity watermarking - applying informed coding and embedding. Mémoire de D.E.A., Institut Eurécom et NEC Research Institute Inc., 2001.

[EBG01] Joachim Eggers, Robert Bäuml et Bernd Girod : Digital watermarking facing attacks by amplitude scaling and additive white noise. Dans *4th International ITG Conference on Source and Channel Coding*, pages 28–30, Berlin, Germany, january 2001.

[EBTG03] Joachim Eggers, Robert Baüml, Roman Tzschoppe et Bernd Girod : Scalar Costa scheme for information embedding. *IEEE Transactions on Signal Processing*, 51(4):1003–1019, april 2003.

[Fil04] Thomas Fillon : *Traitement Numérique du Signal Acoustique pour une Aide aux Malentendants.* Thèse de doctorat, École Doctorale d'Informatique, Télécommunications et Électronique de Paris, GET - Télécom Paris, Paris, France, décembre 2004.

[Fle87] R. Fletcher : *Practical Methods of Optimization.* John Wiley and Sons, 1987.

[Fur02] Teddy Furon : *Application du tatouage numérique à la protection de copie.* Thèse de doctorat, École Doctorale d'Informatique, Télécommunications et Électronique de Paris, GET - Télécom Paris, Paris, France, mars 2002.

[Gar99] Ricardo Garcia : Digital watermarking of audio signals using psychoacoustic auditory model and spread spectrum theory. Dans *107th Convention of Audio Engineering Society (AES)*, New York, USA, september 1999. Preprint 5073.

[GB00] James Gordy et Trevor Burton : Performance evaluation of digital audio watermarking algorithms. Dans *IEEE Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 157–160, Lansing, MI, USA, august 2000.

[GCG+03] Leandro de C. T. Gomes, Pedroe Cano, Emilia Gomèz, Madeleine Bonnet et Eloi Battle : Audio watermarking and fingerprinting : for which applications ? *Journal of New Music Research*, 32(1):65–81, march 2003.

[GM04] Laurent Girin et Sylvain Marchand : Watermarking of speech signals using the sinusoidal model and frequency modulation of the partials. Dans *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Montréal, Canada, may 2004.

[Gol81] Solomon W. Golomb : *Shift Register Sequences.* Aegean Park Press, Laguna Hills, CA, USA, 1981.

[Gom02] Leandro de C. T. Gomes : *Tatouage de signaux audio.* Thèse de doctorat, Université Paris V, Paris, France, juillet 2002.

[Goy02] Alain Goyé : La perception auditive. Polycopié d'enseignement pour Télécom Paris, 2002.

[GP80] S. J. Gel'fand et Mark S Pinsker : Coding for the channel with random parameters. *Problems of Control and Information Theory*, 9(1):19–31, 1980.

[GPM04] Maria Gkizeli, Dimitris A. Padosy et Michael J. Medley : SINR, bit error rate and shannon capacity optimized spread-spectrum steganography. Dans *IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 1561–1564, Singapour, october 2004.

[Hay96] M. Hayes : *Statistical digital signal processing and modeling.* Wiley, 1996.

[Her02] Cormac Herley : Why watermarking is nonsense. *IEEE Signal Processing Magazine*, 19(5):10–11, september 2002.

[HHKK04] Martin Hagmüller, Horst Hering, Andreas Kröpfl et Gernot Kubin : Speech watermarking for air traffic control. Dans *12th European Signal Processing Conference (EUSIPCO)*, pages 1653–1656, Vienna, Austria, september 2004.

[HS02] Ching-Tang Hsieh et Pei-Ying Sou : Performance evaluation of digital audio watermarking algorithms. Dans *IEEE 14th International Conference on Digital Signal Processing (DSP)*, volume 2, pages 705–708, Santorini, Grece, july 2002.

[KA03] Darko Kirovski et Hagai Attias : Audio watermark robustness to desynchronization via beat detection. Dans *5th International Workshop on Information Hiding (IH)*, volume 2002, pages 160–176, Noordwijkerhout, The Netherlands, october 2003.

[Kim03] Hyoung Joong Kim : Audio watermarking techniques. Dans *Pacific Rim Workshop on Digital Steganography*, Kitakyushu, Japan, july 2003.

[KM01] Darko Kirovski et Henrique Malvar : Spread-spectrum audio watermarking : Requirements, applications, limitations. *IEEE International Workshop on Multimedia Signal Processing*, 51(4):219–224, may 2001.

[KM03] Darko Kirovski et Henrique Malvar : Spread-spectrum watermarking of audio signals. *IEEE Transactions on Signal Processing*, 51(4):1020–1033, april 2003.

[KP99] Martin Kutter et Fabien Petitcolas : A fair benchmark for image watermarking systems. Dans *SPIE Electronic Imaging : Security and Watermarking of Multimedia Content*, volume 3657, pages 226–239, San Jose, California, USA, january 1999.

[Lan] Andreas Lang : Stirmark benchmark for audio (smba) : Evaluation of watermarking schemes for audio. http ://amsl-smb.cs.uni-magdeburg.de/smfa/allgemeines.php.

[Lar05] Sonia Larbi : *Structures d'égalisation en tatouage audionumérique*. Thèse de doctorat, Ecole Nationale d'Ingénieurs de Tunis (ENIT) et Télécom Paris (ENST), mars 2005.

[LDC04] Miller Matthew L., Gwenal Doerr et Ingemar Cox : Applying informed coding and embedding to design a robust high-capacity watermark. *IEEE Transactions on Signal Processing*, 6:792–807, June 2004.

[LFT02] Rosa Lancini, Mapelli Francesco et S. Tubaro : Embedding indexing information in audio signal using watermarking technique. Dans *IEEE Region and international symposium on video/image processing and multimedia communications (VIPromCom)*, pages 257–261, Zadar, Croatia, june 2002.

[LGP04] Alejandro Lobo Guerrero Perez : *Etude de techniques de tatouage audio pour la transmission de données*. Thèse de doctorat, Institut National Polytechnique de Grenoble (INPG), Grenoble, France, Décembre 2004.

[LGPD03] Gaëtan Le Guelvouit, Stéphane Pateux et Jonathan Delhumeau : Construction de codes pour tatouage avec prise en compte de l'information adjacente. Dans *GRETSI*, Paris, France, septembre 2003.

[LH00] Sang-Kwang Lee et Yo-Sung Ho : Digital audio watermarking in the cepstrum domain. *IEEE Transactions on Consumer Electronics*, 46(3):744–750, august 2000.

[LMBL04] Alejandro LoboGuerrero, Ferran Marques, Patrick Bas et Joel Lienard : Enhanced audio data synchronization using non linear filters. Dans *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 2, pages 885–888, Montreal, Canada, may 2004.

[LYS05] Zhe-Ming Lu, Bin Yan et Sheng-He Sun : Watermarking combined with celp speech coding for authentication. *IEICE Transactions on Information and Systems*, E88-D(2):330–334, february 2005.

[MB99] Matthew L. Miller et Jeffrey A. Bloom : Computing the probability of false watermark detection. Dans *International Workshop on Information Hiding*, Dresden, Germany, septembre-octobre 1999.

[MCB00] Matthew Miller, Ingemar J. Cox et Jeffrey Bloom : Informed embedding : exploiting image and detector information during watermark insertion. Dans *IEEE International Conference on Image Processing (ICIP)*, pages 1–4, Vancouver, Canada, september 2000.

[McE87] Robert McEliece : *Finite field for scientists and engineers*. Kluwer Academic Publishers, 1987.

[MDC02] Matthew Miller, Gwenaël Doërr et Ingemar J. Cox : Dirty-paper treillis codes for watermarking. Dans *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 129–132, Rochester, NY, USA, september 2002.

[MF03] Henrique Malvar et Darko Florencio : Improved spread spectrum : a new modulation technique for robust watermarking. *IEEE Transactions on Signal Processing*, 51(4):898–905, april 2003.

[MGN02] Tiberiu Muntean, Eric Grivel et Mohamed Najim : Audio digital watermarking based on hybrid spread spectrum. Dans *Second International Conference on Web Delivering of Music (WEDELMUSIC)*, pages 150–155, Darmstadt, Germany, december 2002.

[MHSF03] Michael Mullarkey, Neil Hurley, Guenole Silvestre et Teddy Furon : Application of side-informed embedding and polynomial detection to audio watermarking. Dans *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 57–60, Hong Kong, april 2003.

[Mil01] Matthew Miller : Watermarking with dirty-paper codes. Dans *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 538–541, Thessaloniki, Greece, october 2001.

[Mou05] Ralf Moulin, Pierre et Koetter : Data-hiding codes. *to appear in Proceedings IEEE*, december 2005.

[MS04] Joceli Mayer et Rafael Araujo Silva : Efficient informed embeddinf of multi-bit watermark. Dans *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 3, pages 399–402, Montréal, Canada, may 2004.

[MT01] Mohamed Mansour et Ahmed Tewfik : Audio watermarking by time-scale modification. Dans *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 3, pages 1353–1356, Salt Lake City, UT, USA, may 2001.

[Nor93] Norme internationale ISO/CEI 11172 : Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1.5 Mbits/s, 1993.

[NTK02] Taiga Nakamura, Ryuki Tachibana et Seiji Kobayashi : Automatic music monitoring and boundary detection for broadcast using audio watermarking. Dans *SPIE Electronic Imaging : Security and Watermarking of Multimedia Content IV*, volume 4675, pages 170–180, San Jose, California, USA, january 2002.

[Pet01] Rade Petrovic : Audio signal watermarking based on replica modulation. Dans *5th International Conference on Telecommunications in Modern Satellite, Cable, and Broadcasting Service (TELSIKS)*, volume 1, pages 227–234, Nis, Yugoslavia, september 2001.

[Phi] Philips : Content identification. http ://www.research.philips.com/initiatives/contentid/index.html.

[PLG03] Stéphane Pateux et Gaëtan Le Guelvouit : Performances d'un système de tatouage soumis à des désynchronisations. Dans *GRETSI*, Paris, France, septembre 2003.

[PMMeS92] B. Paillard, P. Mabilleau, S. Morisette et J. et Soumagne : Perceval : Perceptual evaluation of the quality of audio signals. *Journal of Audio Engineering Society*, 40:21–31, march 1992.

[Pop95] Andrei Popescu : *Codage de parole CECP à excitation vectorielle codée par treillis.* Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France, 1995.

[Pro01] John Proakis : *Digital communications.* McGraw-Hill, New York, USA, fourth édition, 2001.

[PS00] Ted Painter et Andreas Spanias : Perceptual coding of digital audio. *Proceedings of the IEEE*, 88(4):451–515, april 2000.

[RJ93] Lawrence Rabiner et B. H. Juang : *Fundamentals of Speech Recognition.* New Jersey : Prentice Hall, april 1993.

[RJ01] M. Rudolf et B. Jechoux : Design of concatenated extended complementary sequences for inter-base station synchronization in wcdma tdd mode. Dans *IEEE Global Telecommunications Conference (GLOBECOM)*, volume 1, pages 674–679, San Antonio, TX, USA, november 2001.

[Rés01] Réseau National des Télécommunications (RNRT) : Aquaflux : Aqua marquage de flux multimédia (audio et vidéo) pour leur transmission, diffusion, circulation, distribution sur les réseaux hétérogènes de télécommunication. http ://www.telecom.gouv.fr/rnrt/rnrt/projets/res_36_ap00.htm, 2001.

[Rés03] Réseau National des Télécommunications (RNRT) : Artus : Animation d'un codeur de langage parlé complété virtuel par tatouage audiovisuel dans un service d'aide à la compréhension des sourds et mal-entendants en télédiffusion numérique. http ://www.telecom.gouv.fr/rnrt/rnrt/projets/res_01_37.htm, 2003.

[SDM] SDMI : The secure digital music initiative. http ://www.sdmi.org.

[SFK02] Shinya Saito, Toshihiro Furukawa et Katsumi Konishi : A data hiding for audio using band division based on QMF bank. Dans *IEEE International Symposium on Circuits and Systems (ISCAS)*, volume 3, pages 635–638, Scottsdale, AR, USA, may 2002.

[Sha48] Claude Shannon : A mathematical theory of communication. *Bell System Technical Journal*, 27(2):379–423, july 1948.

[Sha58] Claude Shannon : Channel with side information at the transmitter. *IBM Journal of Research and Development*, (2):222–293, october 1958.

[SM04] Ariel Sagi et David Malah : Data embedding in speech signals using perceptual masking. Dans *12th European Signal Processing Conference (EUSIPCO)*, pages 1657–1660, Vienna, Austria, september 2004.

[SNHK01] Frank Siebenhaar, Christian Neubauer, Jurgen Herre et Ralph Kulessa : New results on combined audio compression/watermarking. Dans *111th Convention ofAudio Engineering Society (AES)*, New York, USA, november-december 2001. Preprint 5442.

[Spo97] Thomas Sporer : Objective audio signal evaluation - applied psychoacoustics for modeling the perceived quality of digital audio. Dans *103rd Audio Engineering Society Convention (AES)*, New York, USA, october 1997. Preprint 4512.

[SPR+01] Martin Steinebach, Fabien Petitcolas, Frédéric Raynal, Jana Dittmann, Caroline Fontaine, Christian Seibel, Nazim Fatès et Lucilla Croce Ferri : Stirmark benchmark : audio watermarking attacks. Dans *IEEE International Conference on Information Technology : Coding and Computing*, pages 49–54, Las Vegas, NV, USA, april 2001.

[Tac03] Ryuki Tachibana : Audio watermarking for live performance. Dans *SPIE Electronic Imaging : Security and Watermarking of Multimedia Content V*, volume 5020, pages 32–43, San Clara, USA, january 2003.

[TB96] John F. Tilki et A. A. Beex : Encoding a hidden digital signature onto an audio signal using psychoacoustic masking. Dans *7th International Conference on Digital Signal Processing Applications and Technology*, pages 476–480, Boston, MA, USA, october 1996.

[Tel] Teletrax : Global television tracking and reporting. http ://www.teletrax.tv.

[TK96] Thilo Thiede et Ernst Kabot : A new perceptual quality measure for bit rate reduced audio. Dans *100th Convention ofAudio Engineering Society (AES)*, Copenhague, Danemark, november 1996. Preprint 4280.

[Ung87] Gottfried Ungerboëck : Treillis coded modulation with redundant signal set. Part. I : Introduction. Part. II : State of the art. *IEEE Transactions on Communications*, 25(1):5–21, february 1987.

[Uni01] Union Internationale des Télécommunications (UIT) : Recommandation B.S. 1387 : Méthode de mesure objective de la qualité du son perçu, 2001.

[Vit95] Andrew J. Viterbi : *CDMA : Principles of Spread Spectrum Communication*. Addison Wesley Publishing, may 1995.

[WHHS04] Shaoquan Wu, Jiwu Huang, Daren Huang et Y.-Q. Shi : Self-synchronized audio watermark in DWT domain. Dans *IEEE International Symposium on Circuits and Systems (ISCAS)*, volume 5, pages 23–26, Vancouver, Canada, may 2004.

[WHHS05] Shaoquan Wu, Jiwu Huang, Daren Huang et Y.Q. Shi : Efficiently self-synchronized audio watermarking for assured audio data transmission. *IEEE Transactions on Broadcasting*, 51(1):69–76, March 2005.

[ZBD04] Abdellatif Zaidi, Rémy Boyer et Pierre Duhamel : A scaled plus noise model for digital watermarking. application to time jitter. Dans *IEEE Asilomar Conference on Signals, Systems and Computers*, pages 124–135, Pacific Grove, California, USA, june 2004.

[ZBDre] Abdellatif Zaidi, Rémy Boyer et Pierre Duhamel : Audio watermarking under desynchronization and additive noise attacks. *IEEE Transactions on Signal Processing*, (A paraître).

[ZF81] E. Zwicker et R. Feldtkeller : *Psychoacoustique - L'oreille récepteur d'information*. Masson, 1981.

[Zöl03] Udo Zölzer, éditeur. *DAFX - Digital audio effects*. John Wiley and Sons, Ltd, 2003.

[ZP01] Rodger E. Ziemer et Roger L. Peterson : *Introduction to digital communication*. Prentice-Hall, Upper Saddle River, NJ, USA, second edition édition, 2001.

Publications

Article de journal

> Cléo Baras, Nicolas Moreau et Pzremyslaw Dymarski : Controlling the inaudibility and maximizing the robustness in an audio data hiding system. Accepté à *IEEE Transactions on Speech and Audio Processing*, 2005.

Conférences internationales

> Cléo Baras, Nicolas Moreau et Pzremyslaw Dymarski : Comparative study of two informed embedding strategies for audio spread-spectrum data hiding systems. Dans *European Signal Processing Conference (EUSIPCO)*, Antalya, Turquie, september 2005.

> Cléo Baras et Nicolas Moreau : An audio spread-spectrum data hiding system with an embedding strategy adapted to a wiener filtering based receiver. Dans *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1022-1025, Amsterdam, Holland, july 2005.

> Cléo Baras, Pzremyslaw Dymarski et Nicolas Moreau : An audio watermarking scheme based on an embedding strategy with maximized robustness to perturbations. Dans *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 4, pages 357-360, Montreal, Canada, may 2004.

Workshops

> Gérard Bailly, Cléo Baras, Patrick Bas, Séverine Baudry, Rémi Brun, Denis Beautemps, Franck Davoine, Frédéric Elisei, Guillaume Gibert, Denis Grison, Jean-Pierre Léoni, Joël Liénard, Nicolas Moreau, Philippe Nguyen, "ARTUS : calcul et tatouage audiovisuel des mouvements d'un personnage animé virtuel pour l'accessibilité d'émissions télévisuelles aux téléspectateurs sourds comprenant la Langue Française Parlée Complétée", accepté à *Handicap*, 2006

> Cléo Baras et Nicolas Moreau : Modulation CDMA informée ; dans un système de tatouage audio. Dans *Compression et Représentation des Signaux Audiovisuels (CORESA)*, Rennes, France, novembre 2005. A paraître.

> Cléo Baras, Nicolas Moreau et Bassem Zayen, Mécanisme de synchronisation dans des

systèmes de tatouage audio pour des perturbations désynchronisantes à forte dérive. Dans *GRETSI*, Louvain-la-Neuve, Belgique, septembre 2005.

> Pzremyslaw Dymarski, Cléo Baras et Nicolas Moreau : Znakowanie wodne sygnalów dzwiekowych w petli zamknietej ("Approche en boucle fermée du problème de tatouage audio"). Dans *Congrès polonais des Télécommunications*, volume A, pages 221-231, Bydgoszcz, Pologne, septembre 2004.

> Cléo Baras, Pzremyslaw Dymarski et Nicolas Moreau : Système de tatouage en boucle fermée. Dans *GRETSI*, Paris, France, septembre 2003.

> Pzremyslaw Dymarski, Cléo Baras et Nicolas Moreau : Algorytmy znakowania wodnego sygnalów dzwiekowych ("Algorithmes pour le tatouage de signaux audio"). Dans *Congrès polonais des Télécommunications*, volume C, pages 26-34, Bydgoszcz, Pologne, septembre 2003.

> Cléo Baras, Pzremyslaw Dymarski et Nicolas Moreau : Spread spectrum modulations and alternative receiver scheme in an audio watermarking system. Dans *Workshop Cost 276, Transmitting, processing and watermarking multimedia contents*, Lyon, France, mars-avril 2003.

> Cléo Baras, Nicolas Moreau : Modulations par étalement de spectre dans un système de tatouage audio. Dans *Compression et Représentation des Signaux Audiovisuels (CORESA)*, Lyon, France, janvier 2003.

Conférences sans actes

> Cléo Baras, Pzremyslaw Dymarski et Nicolas Moreau : Stratégies d'insertion dans un système de tatouage audio. Dans *Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal audio (JJCASS)*, Marseille, France, mars, 2005.

Rapports techniques

> Cléo Baras, Nicolas Moreau, Alejandro Lobo Guerrero et Patrick Bas : Procédés d'insertion audio basés sur le tatouage. Dans *Rapport d'avancement du projet ARTUS*, octobre, 2003.