# Tatouage de Signaux de Musique
# (Méthodes de synchronisation)

Emilia Gomez

DEA ATIAM

Université de la Méditerranée

Ecole Nationale Supérieure de Télécommunications

Encadré par Nicolas Moreau

27 juillet 2000

# Table des matières

## Remerciements

Je tiens tout d'abord à remercier mon responsable de stage, Nicolas Moreau, pour m'avoir si bien encadré pendant mon stage. J'ai beaucoup appris grâce à sa façon de travailler et j'ai pu profiter de ses conseils.

Je remercie aussi le département TSI de l'ENST pour m'avoir accueillie pendant ces quatre mois de stage, aux chercheurs enseignants qui ont répondu à mes questions, et aux stagiaires et tésards. Je remercie spécialement Rima Hleiss pour m'avoir prété sa machine pendant son absence, et Vincent, Yohan, Anaid, Hichem, Caroline, Alexandre, Thomas, Jamal et Pierre, pour les bons moments passés à l'ENST.

Merci aussi à Leandro de Campos Teixeira Gomes pour l'intêret qu'il a montré pour mon travail et pour ses conseils.

Merci à mes deux collocataires de bureau, Elyane et Cathérine, pour leur soutien et leur bonne humeur.

Grâce à toutes cettes personnes, cette expérience d'initiation à la recherche a été très enrichissante.

# Présentation de l'ENST

L'Ecole Nationale Supérieure des Télécommunications appartient au groupe des écoles des télécommunications, établissement public administratif qui comporte également l'Ecole Nationale Supérieure des Télécommunications de Bretagne et l'Institut National des Télécommunications.

Elle appartient également au groupe des écoles d'ingénieurs de Paris (GEI Paris) qui rassemble les huit autres écoles suivantes : Ecole Nationale Supérieure d'Agronomie, Arts et Métiers, Ecole de Chimie de Paris, Ecole des Eaux et Forêts, Ecole du Génie Rural, Ecole des Mines, Ecole de Physique Chimie de Paris, Ecole des Ponts et Techniques Avancées.

Il y quatre départements d'Enseignement et de Recherche :

- communications et électronique,
- économie, gestion, sciences sociales et humaines ;
- traitement du signal et des images ;
- informatique et réseaux.

## Le département TSI

Le département TSI a pour missions l'enseignement, la recherche, et la formation par la recherche, dans les domaines du traitement du signal et des images et de leurs applications dans divers contextes de la société de l'information, dont les télécommunications.

Ses missions se déclinent comme suit :

- la formation de base : il s'agit de fournir à tous les élèves de l'ENST la maîtrise de certaines notions de base de mathématiques, de probabilités et de traitement du signal et des images ;
- la formation spécialisée afin que certains élèves appréhendent, à divers niveaux d'expertise, les concepts qui leur permettront de traiter les signaux et les images et de participer activement aux secteurs de l'industrie et de la recherche qui les mettent en œuvre ;
- la formation par la recherche pour laquelle le département TSI est fournisseur à la fois de formation (enseignements de DEA, enseignements doctoraux, séminaires

départementaux) et de prestation de recherche (auprès des industriels ou des organismes nationaux et internationaux promoteurs de la recherche). Ces enseignements renforcent l'insertion de TSI dans la communauté universitaire de l'enseignement supérieur ;

- la recherche méthodologique et fondamentale, en relation étroite avec les organismes nationaux et internationaux de coordination de la recherche et en particulier le CNRS ; elle permet au département de contribuer à l'innovation par la découverte de concepts nouveaux ;

- la recherche appliquée, souvent menée en collaboration avec des partenaires industriels français ou étrangers, elle garantit un contact permanent avec les technologies émergentes ainsi qu'avec les nouveaux usages.

Le département TSI participe au rayonnement de l'ENST en la représentant, dans son domaine d'activité, auprès des différentes instances et des organismes nationaux ou internationaux (CNRS, IEEE, RNRT, etc.) et en participant de façon active à la vie scientifique nationale.

## L'organisation du département TSI

Le département TSI est organisé en 5 groupes. Certains groupes sont plus orientés vers la recherche académique, d'autres vers la recherche appliquée, d'autres enfin vers l'enseignement. Chacun de ces groupes contribue cependant à l'ensemble des missions du département.

- groupe *Traitement et Interprétation des images* : ce groupe conduit des recherches sur la mise en œuvre de schémas complets de traitement, d'analyse et d'interprétation d'images, en particulier de scènes complexes. Les domaines d'application des méthodes théoriques qui y sont développées concernent en particulier :

  - l'imagerie médicale et notamment l'imagerie cérébrale anatomique et fonctionnelle ;
  - l'imagerie aérienne et satellitale avec un intérêt particulier sur l'imagerie radar et l'imagerie aérienne à très haute résolution des milieux urbains ; et
  - a description des objets complexes tridimensionnels pour leur analyse ou leur rep résentation fixe ou animée.

- groupe *Traitements statistiques et applications aux communications* : ce groupe travaille dans les domaines suivants :

  - le signal pour les communications ;
  - la séparation de sources ;

  - la modélisation statistique pour le signal et l'image pour laquelle on peut mettre en évidence deux voies de recherche : la reconstruction et la restauration d'images, le télétraffic (analyse et modélisation).

- groupe *Perception, Apprentissage et Modélisation* : ce groupe étudie le rôle des facteurs humains dans l'accès aux divers types d'information :

  - la parole : reconnaissance et identification de locuteurs ;
  - l'image : psychovision (perception du contraste, de la couleur, du relief), et imagerie de très haute qualité ;
  - l'écrit : fax, structuration des documents ;
  - la fusion des modalités perceptives dans l'appréhension de l'environnement ;
  - les interfaces multimodales.

- groupe *Codage* : ce groupe travaille sur des techniques éprouvées de compression de sources ainsi que sur leur adaptation aux applications de l'audiovisuel et du multimédia :

  - la compression de parole et de son ;
  - le codage d'images ;
  - la transmission audiovisuelle sur réseau ;
  - les systèmes temps réel et l'adaptation source-canal ;

- groupe *Audio, Acoustique et Ondes* : ce groupe étudie la physique des ondes dans les deux domaines de l'optique et de l'acoustique.

  - en acoustique : modélisation de la production des sons, perception (psychoacoustique et prothèses auditives) et antennes acoustiques ;
  - en optique : stockage de l'information dans les milieux optiques réinscriptibles.

# Chapitre 1

# Introduction

## 1.1 Nécessité

Les conséquences du développement de la distribution de musique sur internet sont importantes. Ainsi, de nombreux titres sont déjà disponibles sur l'internet, souvent légalement, parfois sans autorisation. Les prochaines années s'annoncent donc comme celles d'un tournant dans l'histoire discographique, d'un tremblement dans les fondements de l'industrie musicale. Le format de compression audio MP3 (MPEG-1 Audio Layer 3) a été défini avec l'objectif principal d'atteindre un taux de compression important des fichiers son sans que la qualité sonore soit altérée. En effet, la différence entre le son original d'un CD et le son de ce même CD qu'on a compressé en MP3 est inaudible et les fichiers obtenus sont de taille tout à fait raisonnable : moins de 1 Mo pour une minute de son qualité CD.

De nombreux chanteurs se sont emparés d'internet pour créer des communautés d'admirateurs dont ils peuvent, en retour, mieux connaître les désirs. Tous ces sites donnent des nouvelles de l'artiste, retracent son histoire, présentent sa discographie et bien sûr mettent des chansons et des vidéos en ligne. Ce sont avant tout des instruments de promotion pour les disques. Parfois, les sites des artistes offrent des titres avant leur sortie dans le commerce et la retransmission en direct de concerts est devenue une habitude. Certains artistes ne voient pas forcément mal le fait qu'internet devienne une véritable base de données de fichiers MP3 illégaux. Nombreux estiment en effet que le MP3 est une nouvelle et extraordinaire manière de faire connaître leurs oeuvres, d'autant plus qu'à l'aide d'Internet, ils peuvent toucher un public plus large.

Internet modifie profondément l'équilibre économique établi par les éditeurs en faveur des artistes. Aujourd'hui, aux Etats-Unis, la diffusion en ligne représente environ cent quarante millions de dollars, soit à peine 1.1% du chiffre d'affaires de l'industrie musicale (voir [18]). Avec l'arrivée des techniques récentes précédemment citées, le risque est de voir des disques uniquement commercialisés par l'intermédiaire du Web et même de voir des artistes court-circuiter leurs maisons de production en produisant et en distribuant eux-mêmes leur musique. Pour distribuer leur musique, il leur faudra seulement répondre à la demande, ce qui leur permettra d'échapper à l'emprise des majors. Les œuvres musicales sont de plus en plus présentes sur internet et il est clair qu'à plus ou moins long terme les supports classiques feront place aux produits numériques. Pour le producteur et le consommateur, le passage de la vente d'albums à celle de morceaux de musique dématérialisés peut représenter une économie notable.

D'un point de vue légal (voir [18]), c'est lors de l'enregistrement d'un fichier son sur un ordinateur que se pose un problème de copyright. En effet, la loi autorise un nombre infini de copies de musique à partir du moment où on s'est acquitté des droits d'auteur, c'est-à-dire si on posséde le CD original. Elle précise également que toutes les copies doivent être à usage privé. C'est pour cette raison qu'on rentre dans l'illégalité lorsqu'on télécharge gratuitement des fichiers MP3 proposés sur internet, à partir du moment où on ne possède pas l'original.

Il est clair qu'il est de plus en plus nécessaire de trouver un format audio sécurisé pour contrôler les recopies frauduleuses et ainsi pouvoir protéger les droits d'auteur, en définissant des nouvelles méthodes de commercialisation et tarification.

Ce problème ne peut pas être entièrement résolu par des techniques de cryptographie. Le concept de tatouage (*watermarking*) est utilisé depuis plusieurs années dans le domaine des images. Son utilisation est plus récente dans le domaine de signaux audio. L'idée de base est d'introduire une information particulière (une signature, une information de copyright, un code, etc.), dans le signal transmis. Ce code doit être transparent à l'utilisateur (inaudible) et résistant aux attaques de plusieurs types.

En première analyse, le problème du tatouage peut se ramener à deux questions essentielles. D'abord, ayant acquis un signal par exemple sur Internet, il s'agit de savoir si le signal a été tatoué ou pas. Ensuite, si le signal est tatoué, on peut tenter d'extraire l'information qui a été caché.

## 1.2 Caractéristiques du signal de tatouage

Dans les signaux audio, le signal de tatouage doit vérifier certaines propriétés. Il doit être :

- Inaudible : l'information d'audio doit rester la même pour l'auditeur avant et après l'insertion du signal de tatouage. L'auditeur ne doit pas entendre de perte de qualité dans les signaux tatoués.
- Résistant aux traitements licites qui peuvent être effectués sur le signal : transmission dans le réseau, compression, copie, changement de support, etc.
- Résistant aux traitements illicites ou attaques de la part des pirates.

Le signal tatoué peut subir des opérations licites ou non (compression, changement de fréquence d'échantillonnage, perte d'échantillons) qui font que le propriétaire du signal perd la capacité de retrouver l'information de tatouage. Parmi ces attaques, il y a un problème qui a été isolé : la perte de synchronisme. Il ne s'agit pas simplement d'un décalage de l'axe temporel, il s'agit d'un phénomène plus complexe de perte de référence.

En première partie, il faut faire une modélisation des différentes attaques qui peuvent être effectués sur le signal, pour pouvoir ainsi les simuler dans le système choisi. Après, à l'aide de différentes techniques, il faut trouver la manière de les détecter et de les contrer.

## 1.3 Etat de l'art

Le premier article consacré au tatouage audio a été publié en 1996 par L. Boney, A. H. Tewfik et K. N. Hamdy [3]. Depuis, une dizaine de papiers ont été publiés dans ce domaine (ce qui est très peu relativement au tatouage d'image). Dans cette section, on présente un historique des différents systèmes de tatouage de signaux de musique. Comme il est très vraisemblable que beaucoup de travaux consacrés au tatouage audio ne sont pas publiés pour des raisons de confidentialité, on présentera également des groupes de travail consacrés à cette étude. Cette section est une reprise de l'étude faite dans [12].

### 1.3.1 Articles

Le schéma de tatouage proposé dans [3] est représenté à la figure 1.1. Le processus d'in-

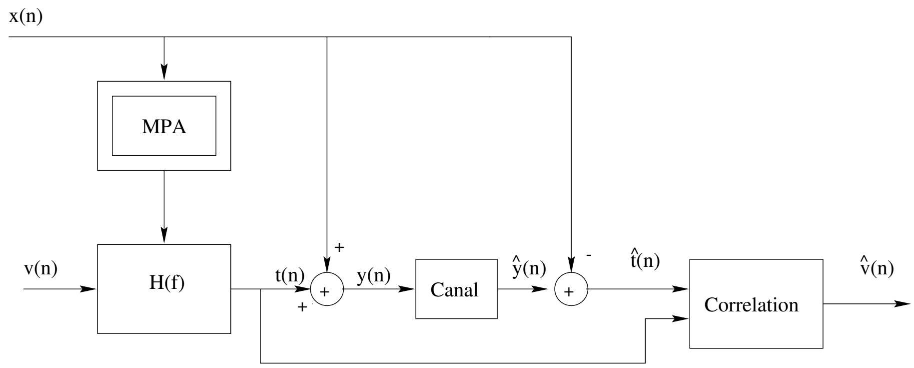


FIG. 1.1 – *Schéma de tatouage proposé dans [3]*

sertion du signal de tatouage à l'émetteur consiste à générer une séquence de bruit pseudoaléatoire $v(n)$ de longueur $N$. Cette séquence est ensuite filtrée par un filtre de réponse en fréquence $H(f)$ qui exploite un seuil de masquage issu d'un modèle psychoacoustique. Pour obtenir un signal de tatouage le plus puissant possible tout en respectant la contrainte d'inaudibilité, le signal de tatouage est mis en forme spectralement en cherchant à faire coïncider la densité spectrale de puissance du tatouage avec le seuil de masquage. Ensuite, la séquence filtrée $t(n)$ est additionnée au signal de musique original $x(n)$. Ces opérations, spécifiques d'une fenêtre d'analyse d'une dizaine de millisecondes environ, sont répétées indéfiniment (jusqu'à la fin du morceau de musique). Le signal tatoué $y(n)$ est transporté dans un canal où il va être soumis vraisemblablement à une opération de compression/décompression simulable par l'addition d'un bruit très coloré additif. Au récepteur, la détection du tatouage tatoué et bruité $\hat{y}(n)$ est réalisée par une opération de corrélation entre $t(n)$ et $\hat{t}(n)$ (une estimée du tatouage). Dans l'étude [3], c'est le propriétaire du signal qui réalise le test : le signal original est supposé connu au récepteur. Il est donc possible de construire $\hat{t}(n)$ en soustrayant $x(n)$ à $\hat{y}(n)$.

La connaissance du signal original au récepteur est une contrainte beaucoup trop forte dans la pratique. Il faut être capable de décider au récepteur si le signal reçu est tatoué ou pas uniquement à partir du signal reçu. Il s'agit d'un problème de test d'hypothèse. Il faut aussi être capable de détecter des données "cachées". C'est un problème de détection. Ce processus de décision peut être représenté par le schéma de la figure 1.2.

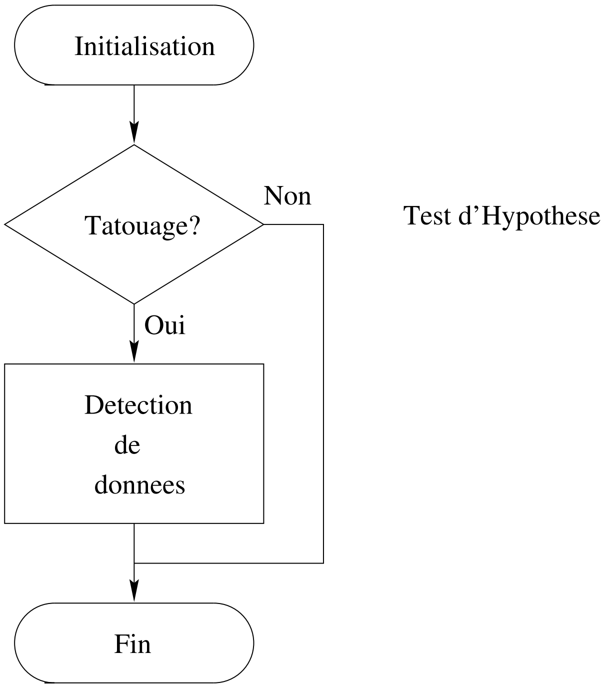


FIG. 1.2 – *Algorithme de détection du tatouage*

T. Furon et P. Duhamel ont introduit récemment [7], [9] une méthode permettant de répondre à la première question sans permettre le deuxième traitement. Le schéma proposé, adapté aux signaux audio, est représenté dans la figure 1.3. L'idée de base est d'imposer des propriétés statistiques particulières au signal de tatouage de telle sorte qu'il soit possible de décider au récepteur si ces propriétés existent ou pas. Une densité spectrale de puissance donnée $S_S(f)$ est imposée au signal $s(n)$. Un entrelaçeur pseudoaléatoire permute les échantillons de chaque fenêtre pour étaler le spectre. Comme dans la méthode précédente, pour augmenter autant que possible la puissance du signal de tatouage tout en

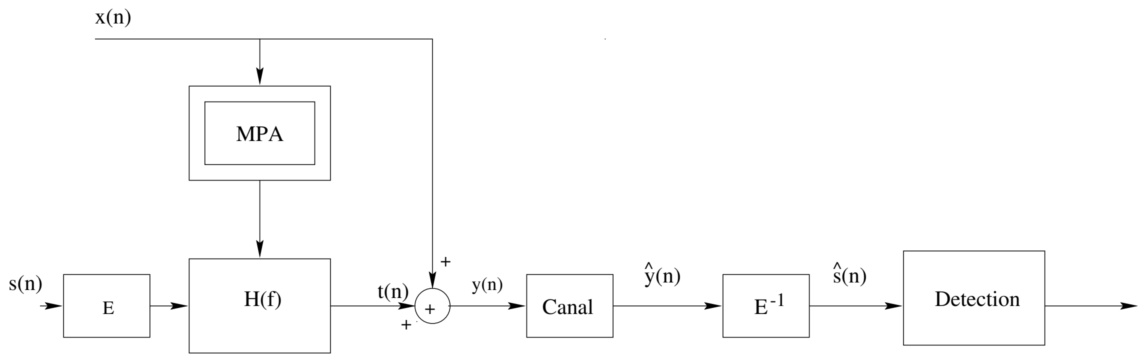


FIG. 1.3 – *Schéma de tatouage proposé dans [7] et [9]*

respectant la contrainte d'inaudibilité, le signal entrelacé est filtré par un filtre de réponse en fréquence déterminé par un seuil de masquage issu d'un modèle psychoacoustique. Avant la détection, il faut appliquer le processus d'entrelacement inverse au signal observé pour blanchir l'influence du signal de musique $x(n)$ et restaurer la composante due au signal de tatouage original $s(n)$. Un test d'hypothèse binaire est ensuite effectué :

- Si le signal n'est pas tatoué (hypothèse $H_0$), le signal observé après l'entrelacement inverse est blanc. Sa densité spectrale de puissance $S_{\hat{S}}(f)$ doit donc être de la forme

$$S_{\hat{S}} = a$$

- Sinon (hypothèse $H_1$), le signal observé possède une densité spectrale de puissance qui est la somme d'une constante et d'une autre composante très fortement colorée due au tatouage. On doit donc avoir :

$$S_{\hat{s}} = a + bS_s(f)$$

Les auteurs ont montré que ce système de tatouage était similaire à un système de cryptographie avec une clé publique et une clé privée. Dans le schéma de tatouage de la figure 1.1, le signal $v(n)$ joue le rôle de clé privée (connue seulement par le propriétaire du signal). Au récepteur, la connaissance de cette clé est nécessaire. Le schéma est alors dit “symétrique”. Dans le schéma de tatouage de la figure 1.3, le propriétaire du signal original peut décider de rendre publique la propriété statistique particulière imposée au signal de tatouage (ici $S_S(f)$) sans donner la suite des échantillons $s(n)$. Au récepteur, seul le test d'hypothèse peut être effectué. Des données peuvent être “cachées”. La connaissance de la clé publique ne permet pas l'extraction du signal de tatouage du signal marqué ce qui assure sa robustesse. Le schéma de tatouage est dit “asymétrique”.

Une approche complètement différente est décrite dans [6]. Dans cette publication, les auteurs définissent une technique de tatouage qui s'applique directement au flux de bits codé par le codeur MPEG-2 AAC. Le tatouage s'effectue en modifiant des facteurs d'échelle dans le flux de bits. Cette technique semble être plus robuste dans le domaine codé AAC mais le tatouage ne perdure pas après le processus de décompression. Une autre technique qui fonctionne dans le domaine AAC a été présentée dans [11]. L'approche est fondamentalement la même qu'en [3] mais le tatouage est inséré directement dans le flux de bits codé. Le tatouage est, cette fois ci, présent après la décompression.

Dans [10], les auteurs introduisent une méthode de cryptographie qui génère, au travers d'une clé donnée (*clé 1*), un signal audio dégradé avec une qualité suffisante à des finalités d'évaluation, permettant à un client potentiel de décider de payer ou non pour la version de haute qualité. Quand le client paie pour cette deuxième version de haute qualité, il reçoit une version cryptée avec une clé (*clé 2*) personnalisée. En utilisant cette clé et un logiciel compatible, le client est capable d'écouter la version de haute qualité mais il n'a pas accès à la version sans cryptage. Le transcodage de la *clé 1* à la *clé 2* est très efficace et peut être effectué en temps réel. Même s'il ne s'agit pas à proprement parler d'une technique de tatouage, elle résout la faiblesse de la majorité des systèmes de protection par cryptage car l'utilisateur n'a pas accès à la version non protégée du signal audio.

Dans [8], l'auteur propose une méthode de tatouage basée sur une modulation BPSK et un processus d'étalement de spectre. A la différence de la technique présentée dans [3], le signal de tatouage n'est pas additionné au signal de musique dans le domaine temporel. Les composantes spectrales du signal audio qui se trouvent en dessous du seuil de masquage sont remplacées par des composantes spectrales du tatouage. Cette approche présente deux inconvénients principaux. D'une part, la puissance du tatouage est faible. D'autre part, la localisation spectrale du tatouage est identifiable par l'estimation du seuil de masquage d'après le signal tatoué : cela pose un problème de robustesse. Par contre, le concept de tatouage audio, vu comme un canal de communication, est clairement présenté dans cet article et est à la base du système de tatouage présenté au chapitre suivant.

### 1.3.2 Groupes de travail

Plusieurs groupes de travail s'intéressent à ce domaine. On peut citer par exemple :

**SDMI** (Secure Digital Music Initiative) ([14])

C'est un forum qui allie plus de 180 sociétés et organisations qui travaillent dans les technologies de l'information, l'électronique, les technologies de la sécurité, l'industrie de la production musicale, et les fournisseurs d'accès Internet. Son but est de développer des spécifications pour protéger la distribution, l'enregistrement et le stockage d'audio numérique pour ce nouveau marché émergeant. Les spécifications des technologies définies par ce groupe de travail sont les suivantes :

- donner à l'utilisateur des accès adaptés à la musique on line et par ce moyen des nouvelles méthodes de distribution numériques,
- permettre la protection des droits d'auteurs,
- promouvoir le développement des nouveaux marchés et technologies dans le domaine de la musique.

**OPIMA** (Open Platform Initiative for Multimedia Access) ([16])
Cette organisation développe des spécifications pour qu'un utilisateur puisse prendre un dispositif récepteur et commencer à consommer et payer pour des services en utilisant un dispositif de contrôle à distance. OPIMA doit donner aux participants la capacité d'acquérir, approvisionner, consommer des services multimédia appartenant à une base mondiale d'accord avec les droit associés avec ces services. OPIMA s'adresse principalement au management et la protection des droits d'auteur. OPIMA est inscrit dans le programme ITA (Industry Technical Agreement) de la Comission International Electrotechnique (IEC).

**MPEG** (Moving Picture Experts Group) ([13])
Il s'agit d'un groupe de travail de l'ISO/IEC chargé du développement de standards de représentation d'audio et vidéo numérique. Après MPEG-1 et MPEG-2, les standards utilisés par la télévision numérique et le DVD et MPEG-4, le standard pour la distribution de multimédia dans le web, MPEG travaille dans MPEG-7, un outil de description multimédia de contenu (*Multimedia Content Description Interface*) et dans un nouveau standard MPEG-21 *"Multimedia Framework"* depuis juin 2000. Dans un monde où les technologies de l'information, l'électronique de consommation et les produits de télécommunications incorporent des technologies de plus en plus raffinées, il y a de plus en plus besoin d'avoir des standards actualisés. MPEG a installé un mécanisme pour inclure les résultats de recherche dans l'industrie. MPEG est un comité de l'ISO/IEC ouvert pour des experts accrédités par un Corps National de Standards. En moyenne, une réunion est organisée pour accueillir 300 experts représentant plus de 200 sociétés dans tous les domaines industriels en rapport avec l'audio et la vidéo numérique, et le multimédia. En moyenne, plus de 20 pays sont représentés.

### 1.3.3 Plan du document

Le document est composé de deux parties, selon les étapes de déroulement du stage :

1. Description du système de tatouage utilisé. Dans cette partie, on donne une description du système de tatouage developpé à l'ENST, son diagramme fonctionnel, ses caractéristiques. Une étude des paramètres de la modulation qu'on peut faire varier pour protéger les données tatouées est également fournie.

2. Méthodes de synchronisation. Cette deuxième partie est la veritable contribution de ce travail d'initiation à la recherche. On verra d'abord comment modéliser les différentes attaques possibles sur le signal tatoué. On fera ensuite une analyse des paramètres permettant de détecter et de contrer ce type d'attaques.

# Chapitre 2

# Schéma de tatouage étudié

## 2.1 Schéma de principe

Considérons le schéma de tatouage présenté dans la figure 2.1. Ce schéma considère le

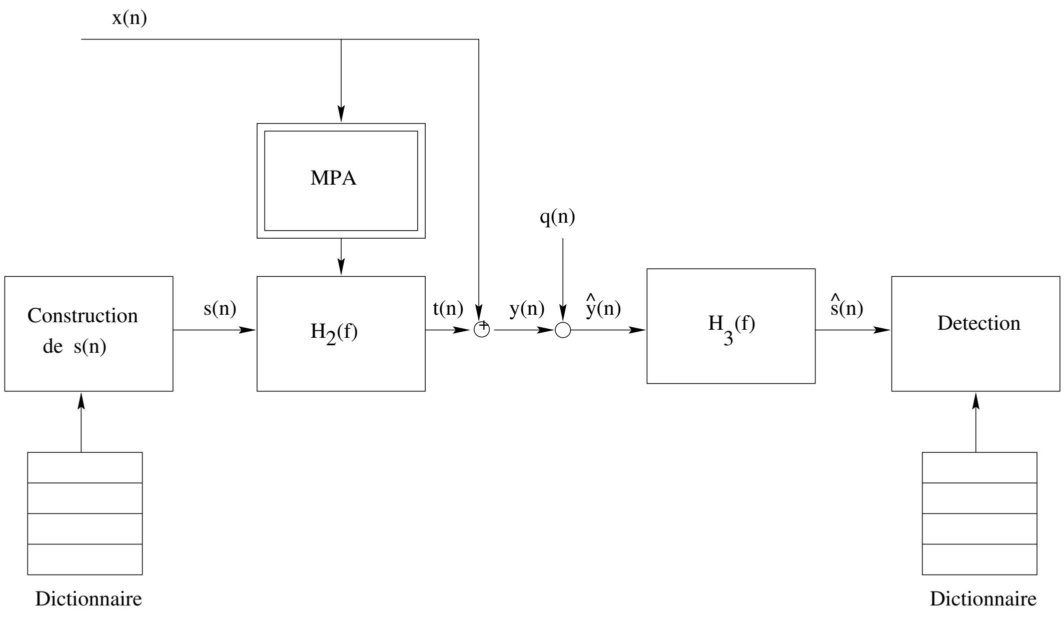


FIG. 2.1 – *Schéma de tatouage proposé*

problème du tatouage comme un problème analogue à celui d'une transmission dans une chaîne de communication très fortement bruitée. Le signal d'information va être affecté par différentes perturbations :

– Bruit dû au signal de musique : le signal de musique peut être considéré comme un bruit très fortement coloré. Sa puissance est très variable au cours du temps.

– Bruit dans le canal de transmission : cette source de bruit est essentiellement due à la compression et peut être modélisée par un bruit gaussien, de moyenne nulle et de densité spectrale de puissance donnée par un seuil de masquage.

– Perturbations dues aux attaques : ce type de perturbation modélise les changements dans la puissance du signal à cause des attaques externes.

Dans ce système de communication, les données à transmettre sont des bits d'information cachés comme signal de tatouage. Le système est donc composé par un transmetteur de l'information de tatouage et un récepteur qui essaie de décoder l'information qui a été envoyée par le transmetteur.

Soit le signal de musique $x(n)$, auquel nous allons additionner le signal de tatouage $t(n)$. Ce signal $t(n)$ est le résultat du filtrage du signal $s(n)$ par un filtre obtenu à partir d'un modèle psychoacoustique qui représente le seuil de masquage associé au signal de musique $x(n)$. Le signal $s(n)$ est composé d'une suite de signaux $s_i(n), n = 0, \ldots, N-1$ appartenant à un dictionnaire de $M$ signaux $\{s_i(n), i = 1, \ldots, M\}$. L'indice $i$ sélectionné tous les $N$ échantillons, dépend du symbole que l'on désire émettre. Le signal tatoué $y(n) = x(n)+t(n)$ est transmis dans un canal. Le récepteur reçoit le signal $\hat{y}(n)$, résultat de la dégradation à travers le canal du signal transmis $y(n)$. Le but est de décoder correctement les données envoyées à partir du signal reçu $\hat{y}(n)$. Au récepteur, on utilise un filtrage adapté suivi d'un banc de corrélateurs qui effectue un calcul de distances et une estimation du symbole transmis.

## 2.2 Construction du dictionnaire

Le signal de tatouage $t(n)$ est généré par une modulation de phase quaternaire d'une porteuse en utilisant la technique d'étalement de spectre (DS/QPSK), comme le montre le schéma de la figure 2.2.

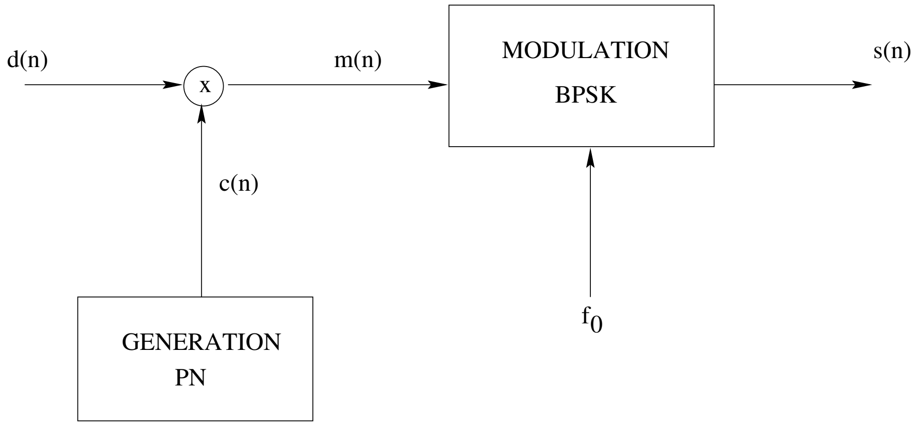


FIG. 2.2 – *Procédure de génération des signaux du dictionnaire*

### 2.2.1 Modulation QPSK

Dans la modulation M-BPSK (*M-ary Bipolar Phase-Shift Keying*), le signal $s(n)$ se construit par concaténation de différents messages. La transmission d'un message parmi $M$ est réalisée par l'envoi d'un signal, un vecteur de dimension $N$, $s_i = [s_i(0), \cdots, s_i(N-1)]^t$. L'ensemble $\{s_i(n)\}$, $i = 1, \ldots, M$ de signaux définit le dictionnaire $S$ employé. Le dictionnaire des signaux est donc composé de $M$ signaux. Si $M = 4$, on a quatre symboles possibles à transmettre, donc 2 bits pour coder ces 4 symboles. Ce type de modulation, 4-BPSK ou QPSK, est employé dans le système de tatouage décrit.

Le dictionnaire $S$ est défini de la manière suivante. Appelons $\{W(i)\}$ la séquence de données binaires du signal de tatouage. On suppose $W(i) = \pm 1$. On cherche à transmettre cette information à un débit de $R_W$ bits par seconde. On choisira par la suite un débit constant $R_W = 125$ bps car, dans la pratique, le but est de transmettre de l'ordre d'une trentaine de bits toutes les secondes : on prend donc une marge de sécurité. Avec une fréquence d'échantillonnage $f_e$ égale à 32 KHz, le temps "bit" $N_b$ est égal à 256 échantillons. Un bit est donc transmis chaque 256 échantillons.

On crée une séquence pseudoaléatoire $W_c$ de longueur $N_c \leq N_b$ afin d'effectuer un étalement de spectre. On réalise ensuite une modulation BPSK en calculant les signaux du dictionnaire :

$$s_i(n) = \sqrt{2}\sin[2\pi f_0 n + W(i)\frac{\pi}{2}] = \sqrt{2}W(i)\cos(2\pi f_0 n) \tag{2.1}$$

La fréquence $f_0$ (normalisée) est la fréquence de la porteuse. La densité spectrale de puissance du signal $s(n)$ est la somme de deux versions translatées, en $f_0$ et $-f_0$, de la densité spectrale de puissance du signal avant la modulation. Pour éviter le recouvrement de spectre entre les deux lobes principaux, il faut que la fréquence de la porteuse soit comprise entre les valeurs suivantes :

$$\frac{R_{Wp}}{f_e} < f_0 < \frac{1}{2} - \frac{R_{Wp}}{f_e} \tag{2.2}$$

Les signaux $\{s_i(n)\}, i = 1, \ldots, M$ peuvent en général s'écrire comme combinaisons linéaires de fonctions appartenant à une base orthonormée $\{\phi_j\}$, $j = 1, \ldots, M'$ obtenue par la méthode d'orthogonalisation de Gram Schmidt appliquée au dictionnaire $S$ :

$$s_i(n) = \sum_{i=1}^{M'} a_{ij}\phi_j(n) \tag{2.3}$$

avec $M' < M$. Dans le cas d'une modulation QPSK, il y a deux fonctions définissant cette base orthonormée :

$$\begin{aligned} \phi_1(n) &= \sqrt{\tfrac{2}{N}}\cos(2\pi f_0 n) \quad n = 0, \ldots, N-1 \\ \phi_2(n) &= \sqrt{\tfrac{2}{N}}\sin(2\pi f_0 n) \quad n = 0, \ldots, N-1 \end{aligned} \tag{2.4}$$

Cette base orthonormée constitue un système de coordonnées dans lequel on peut représenter les signaux du dictionnaire. Cette représentation définit la constellation de signaux de la modulation. Dans le cas QPSK, l'espace des signaux est bidimensionnel. La constellation de signaux est composée par 4 points correspondants aux 4 signaux du dictionnaire comme le montre le tracé de la figure 2.3.

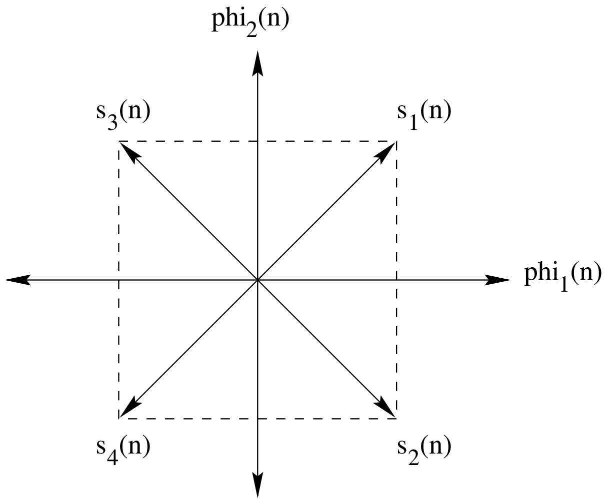


FIG. 2.3 – *Constellation de signaux dans le cas QPSK*

### 2.2.2 Etalement de spectre

Une séquence pseudoaléatoire ( séquence PN) est une suite de valeurs binaires $\{W_c(i)\}, i = 0, \ \ldots, N_c - 1$ avec certaines propriétés d'autocorrélation. L'avantage de l'utilisation de l'étalement de spectre est sa capacité à rejeter des interférences externes. Pour cela, le signal d'information (signal de tatouage) utilise une largeur de bande plus grande que le minimum nécessaire à la transmission. La méthode utilisée pour augmenter la largeur de bande d'un signal d'information consiste à moduler la séquence de données par une séquence pseudoaléatoire $W_c$ de longueur $N_c$. Si la séquence de données a un spectre de largeur de bande $\frac{2}{T_s}$, le signal produit aura un spectre avec une largeur de bande égale à $N_c\frac{2}{T_s}$ qui est la largeur originale multipliée par le facteur $N_c$ comme le montre le tracé de la figure 2.4.

L'étalement de spectre est nécessaire. Si le tatouage se trouve dans une bande de fréquence très étroite, on risque d'avoir une composante de signal de musique très forte dans ces fréquences et le tatouage sera alors totalement bruité. Il faut donc bien étaler pour avoir de l'énergie dans toute la bande utile. La valeur de $N_c$ détermine la largeur de bande du signal en sortie du codeur. Inversement, si on augmente exagérément la longueur $N_c$ de la séquence pseudoaléatoire, on perd la notion de porteuse. Il faut donc arriver à un compromis.

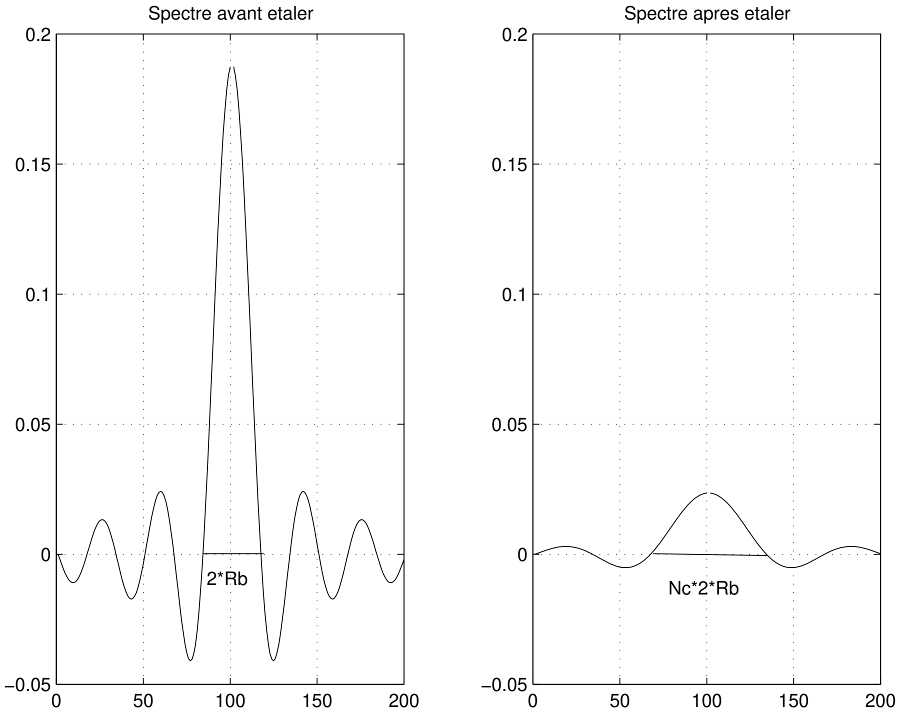


FIG. 2.4 – *Spectres des signaux avant et après la procédure d'étalement de spectre*

## 2.3 Construction du signal de tatouage et du signal tatoué

La première caractéristique du signal de tatouage est qu'il ne soit pas audible pour que le signal tatoué ait la même qualité qu'avant l'addition du tatouage. Pour réussir à respecter cette contrainte, il faut utiliser un modèle psychoacoustique qui étudie le fonctionnement du système auditif humain pour pouvoir déterminer les conditions pour que le signal de tatouage ne soit pas audible. De tous les phénomènes du processus d'audition, nous allons exploiter le phénomène de masquage fréquentiel.

### 2.3.1 Masquage

Le masquage est le phénomène qui produit qu'un son audible devienne inaudible en présence d'un autre son audible, le masqueur. Le phénomène de masquage dépend des caractéristiques temporelles et fréquentielles des deux sons, le son masquant et le son masqué. On a deux types de masquage :

– Le masquage fréquentiel se produit entre deux sons générés en même temps. Quand deux signaux simultanés sont placés à des fréquences proches, le signal le plus puissant rend l'autre signal inaudible. Le seuil de masquage d'un son masquant donné dépend de sa fréquence, de son niveau de pression (*Sound Pressure Level*, SPL) et des caractéristiques, son pur/bruit bande étroite, des signaux masquant et masqué. Par exemple, c'est plus facile qu'un bruit masque un son pur que le contraire, qu'un son pur masque un bruit. Les signaux hautes fréquences sont masqués plus facilement.

- Le masquage temporel peut se diviser en deux phénomènes : le prémasquage et le postmasquage. Les effets de prémasquage rendent inaudibles des signaux faibles produits avant un autre signal plus puissant, le son masqueur. Le phénomène de postmasquage rend inaudible des signaux faibles produits après un signal de forte puissance, le masqueur. Le phénomène de prémasquage se produit entre 5 et 20 ms avant le début du son masqueur et le postmasquage se produit 50-200 ms après la fin du son masqueur.

Dans le système de tatouage, c'est le phénomène de masquage fréquentiel qui est uniquement exploité. En utilisant l'information de masquage fréquentiel, nous pouvons mettre en forme la caractéristique spectrale du signal de tatouage.

### 2.3.2 Modèle psychoacoustique

Le modèle psychoacoustique manipule l'information d'audio pour générer des informations relatives au seuil de masquage final. L'information du seuil de masquage fréquentiel est utilisé pour la mise en forme du signal de tatouage. Le tatouage mis en forme spectralement par le seuil de masquage est imperceptible par l'auditeur moyen (si le modèle psychoacoustique est bien construit). Le seuil de masquage est actualisé toutes les vingtaines de millisecondes environ.

Le modèle psychoacoustique utilisé dans le système de tatouage est le même que celui utilisé par la norme MPEG. Les étapes nécessaires pour générer un modèle psychoacoustique sont représentées dans la figure 2.5. Elles sont les suivantes :

1. Le premier pas consiste à transformer le signal d'audio dans le domaine fréquentiel en utilisant la Transformée de Fourier Rapide (FFT).
2. Dans le domaine fréquenciel, il faut calculer la Densité Spectrale de Puissance du signal, l'énergie et l'énergie étalée par bande critique pour estimer le seuil de masquage.
3. Ce seuil de masquage estimé est utilisé pour mettre en forme spectrale le tatouage.

### 2.3.3 Contrainte d'inaudibilité

Le signal de tatouage sera inaudible si sa densité spectrale de puissance est en dessous du seuil de masquage:

$$S_t(f) \leq S_{masqu}(f) \tag{2.5}$$

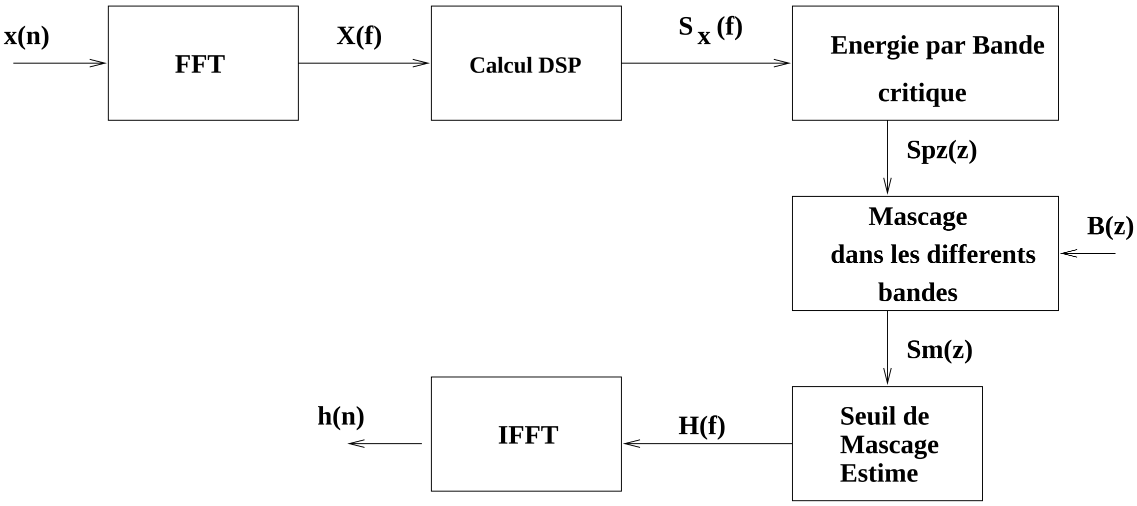


FIG. 2.5 – *Modèle auditif psychoacoustique*

Dans le cas d'un signal $s(n)$ blanc, le filtre de réponse en fréquence $H_2(f)$ peut être synthétisé directement. On exploite la relation

$$\sigma_S^2 |H_2(f)|^2 = S_{masque}(f) \tag{2.6}$$

Avec une modulation BPSK, le signal n'est pas blanc. La figure 2.6 montre le seuil de masquage obtenu à partir de l'analyse du signal de musique $x(n)$. Dans la figure 2.7, on a représenté le spectre du signal de tatouage et le seuil de masquage. Cette figure montre que le spectre du signal de tatouage reste tout de même en dessous du seuil de masquage.

Le signal à transmettre $y(n)$ est le résultat de l'addition échantillon par échantillon du signal de musique $x(n)$ et du signal de tatouage $t(n)$. Au récepteur, on récupère le signal

$$\hat{y}(n) = y(n) + q(n) \tag{2.7}$$

où $q(n)$ modélise l'opération de compression/reconstruction habituellement réalisée en amont et en aval du canl de transmission.

## 2.4 Construction de $\hat{s}(n)$

Au récepteur, on effectue d'abord un filtrage puis un processus de corrélation avec les signaux du dictionnaire pour déterminer les symboles émis. Le filtrage $H_3(f)$ utilisé en réception correspond au filtrage de Wiener, qui est le filtre optimal au sens des moindres carrés.

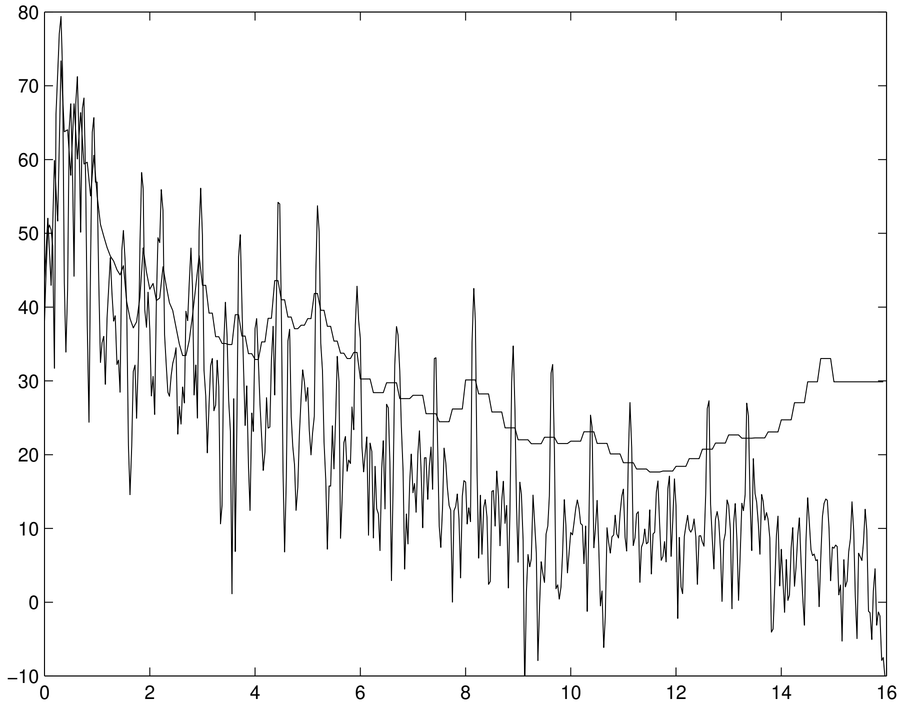


FIG. 2.6 – *Spectre du signal de musique et seuil de masquage*

### 2.4.1 Filtrage de Wiener

Un problème classique d'estimation en théorie des communications est l'estimation linéaire. Il s'agit d'estimer la réponse désirée comme une version filtrée d'un signal d'entrée. Dans cette section, on présente la version à temps discret du filtre de Wiener pour résoudre le problème d'estimation linéaire. On supposera les notations indépendantes de notre problème de tatouage.

Soit $X(n)$ un processus aléatoire stationnaire centré correspondant à un signal bruité. On cherche la réponse impulsionnelle d'un filtre à réponse impulsionnelle finie $h(0), \ldots, h(M)$ telle que la différence entre la réponse souhaitée $D(n)$ et la sortie du filtre $Y(n)$ soit minimale dans un sens statistique. Dans le cas de la prédiction par exemple, $D(n)$ est égal au prochain échantillon $X(n+1)$ de la série temporelle. Soit :

$$\varepsilon(n) = D(n) - Y(n) = D(n) - \sum_{k=0}^{M} h(k)X(n-k) \tag{2.8}$$

l'erreur d'estimation. Cette erreur est une variable aléatoire. Dans la théorie de Wiener, le filtre est optimisé en minimisant la valeur quadratique moyenne de l'erreur d'estimation $\varepsilon(n)$. Soit :

$$\xi = E[\varepsilon^2(n)] \tag{2.9}$$

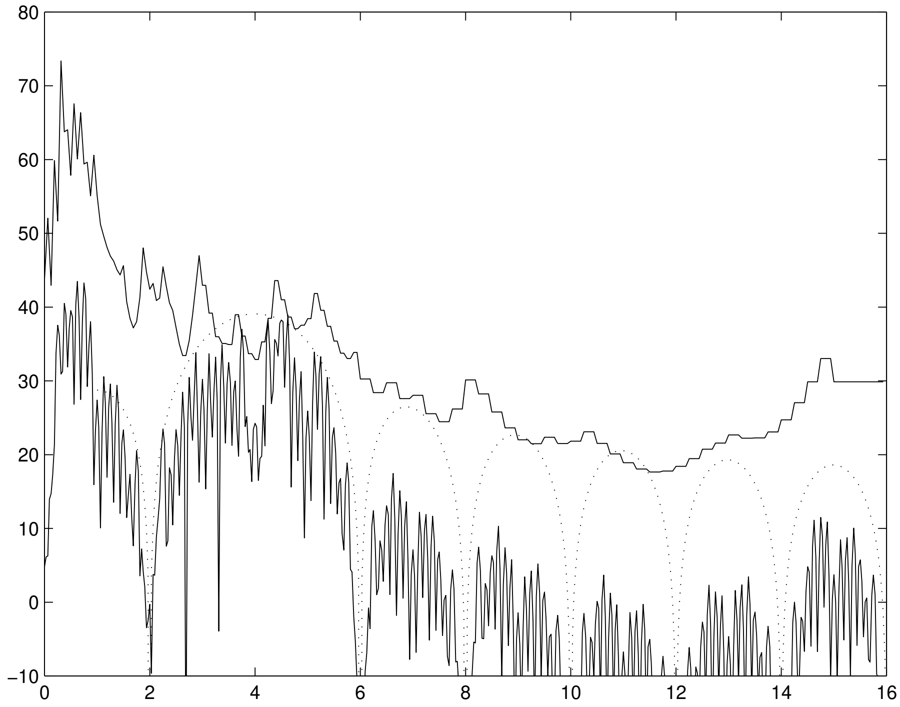


Fig. 2.7 – *Spectre du signal de tatouage et seuil de masquage*

l'erreur quadratique moyenne. On peut écrire

$$\xi = E[D^2(n)] - 2\sum_{k=0}^{M} h(k)E[D(n)X(n-k)] + \sum_{k=0}^{M}\sum_{m=0}^{M} h(k)h(m)E[X(n-k)X(n-m)] \quad (2.10)$$

$$\xi = \sigma_D^2 - 2\sum_{k=0}^{M} h(k)R_{DX}(k) + \sum_{k=0}^{M}\sum_{m=0}^{M} h(k)h(m)R_{XX}(m-k) \quad (2.11)$$

en notant $R_{DX}(k)$ et $R_{XX}(k)$ les fonctions d'intercorrélation et d'autocorrélation. L'erreur quadratique moyenne est une fonction de second ordre des coefficients du filtre. L'erreur quadratique moyenne est minimum quand sa dérivée par rapport aux coefficients du filtre est nulle, c'est à dire:

$$\frac{\partial \xi}{\partial h_k} = -2R_{DX}(k) + 2\sum_{m=0}^{M} h(m)R_X(m-k) = 0 \quad (2.12)$$

Le valeurs optimales des coefficients du filtre $h_{opt}(0), \ldots, h_{opt}(M)$ sont solution des relations de Wiener-Hopf:

$$\sum_{m=0}^{M} h_{opt}(m)R_X(m-k) = R_{DX}(k) \quad k = 0, 1, \ldots, \quad (2.13)$$

L'erreur quadratique moyenne minimale est donnée par :

$$\xi_{min} = \sigma_D^2 - \sum_{k=0}^{M} h_{opt}(k) R_X(k) \tag{2.14}$$

Il n'y a pas d'autre filtre qui puisse produire une erreur quadratique moyenne plus petite que seuil obtenu par le filtre de Wiener calculé précédemment pour des valeurs données des fonctions de corrélation.

## 2.5 Détection

### 2.5.1 Filtrage adapté

Soit le cas de la transmission d'une information binaire : $M = 2$. On recherche le filtre de réponse impulsionelle $h(n)$ ou de réponse en fréquence $H(f)$ et le seuil $K$ avec lequel on va comparer la sortie du filtre observée à l'instant N-1. Ces deux paramètres doivent minimiser la probabilité d'erreur dans la détection. Supposons $y(n)$ l'entrée et $v(n)$ la sortie du filtre $h(n)$. Supposons que le signal $y(n)$ est la somme du signal émis $s_i(n)$ appartenant au dictionnaire et d'un bruit additif $b(n)$. On a

$$v(n) = y(n) * h(n) = [s_i(n) + b(n)] * h(n) \tag{2.15}$$

Etant donné les caractéristiques du bruit (blanc, gaussien, de moyenne nulle et de variance $\sigma_b^2$), l'observation $v(N-1)$ est une variable aléatoire gaussienne avec les caractéristiques suivantes. La moyenne est égale à

$$m_i = E\{v(N-1)\} = \sum_k h(k)[s_i(N-1-k) + E\{b(N-1-k)\}] = \sum_k h(k) s_i(N-1-k) \tag{2.16}$$

et la variance est donnée par

$$\sigma^2 = \sigma_b^2 \int_{-1/2}^{1/2} |H(f)|^2 df \tag{2.17}$$

La densité de probabilité de $v(N-1)$ a donc pour expression :

$$p_V(u) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(u-m_i)^2}{2\sigma^2}) \tag{2.18}$$

La probabilité de faire une erreur est alors donné par l'expression suivante :

$$P_e = \frac{1}{2\sqrt{2\pi\sigma^2}} [\int_K^{+\infty} exp(-\frac{(u-m_0)^2}{2\sigma^2}) du + \int_{-\infty}^{K} exp(-\frac{(u-m_1)^2}{2\sigma^2}) du] \tag{2.19}$$

en supposant les probabilités *a priori* de $s_0(n)$ et $s_1(n)$ équiprobables et $m_0 < m_1$. Les probabilités *a priori* étant égales, on peut aussi dire que la valeur optimale pour le seuil de décision est donnée par :

$$K_{opt} = \frac{m_0 + m_1}{2} \tag{2.20}$$

En appelant $Q(x)$ l'intégrale

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{+\infty} exp(-\frac{x^2}{2})dx \tag{2.21}$$

on obtient finalement

$$P_e = Q(\frac{m_1 - m_0}{2\sigma}) \tag{2.22}$$

Posons $g(n) = s_1(n) - s_0(n)$ et $g_f(n) = g(n) * h(n)$. Minimiser la probabilité d'erreur est équivalent à maximiser

$$\nu = \frac{g_f(N-1)}{\sigma} \tag{2.23}$$

ou

$$\nu^2 = \frac{|g_f(N-1)|^2}{\sigma^2} = \frac{| \int_{-1/2}^{1/2} G(f)H(f)exp(j2\pi f(N-1))df|^2}{\sigma_b^2 \int_{-1/2}^{1/2} |H(f)|^2 df} \tag{2.24}$$

En appliquant l'inégalité de Schwarz:

$$| \int_{-1/2}^{1/2} G(f)H^*(f)df|^2 \leq \int_{-1/2}^{1/2} |G(f)|^2 \int_{-1/2}^{1/2} |H(f)|^2 \tag{2.25}$$

on peut écrire

$$\nu^2 \leq \frac{\int_{-1/2}^{1/2} |G(f)|^2 \int_{-1/2}^{1/2} |H(f)|^2}{\sigma_b^2 \int_{-1/2}^{1/2} |H(f)|^2 df} \tag{2.26}$$

La valeur maximale est atteinte si

$$H(f) = G^*(f)exp(-j2\pi f(N-1)) \tag{2.27}$$

soit

$$h(n) = s_1(N-1-n) - s_0(N-1-n) \tag{2.28}$$

Le récepteur optimale est donc celui qui consiste à passer le signal reçu $y(n)$ à travers de deux filtres dont les réponses impulsionelles sont les versions renversées des deux signaux $s_1(n)$ et $s_0(n)$ puis de comparer leur différence à l'instant $N-1$ à un seuil. La valeur maximale $\nu^2_{max}$ est donnée par

$$\nu^2_{max} = \frac{1}{\sigma_b^2} \sum_n |s_1(n) - s_0(n)|^2 = \frac{1}{\sigma_b^2} [\sum_n |s_0(n)|^2 + \sum_n |s_1(n)|^2 - 2 \sum_n s_0(n)s_1(n)] \tag{2.29}$$

En appelant $E_0$ et $E_1$ les énergies des signaux $s_0(n)$ et $s_1(n)$ que l'on supposera finies, on trouve

$$\nu^2_{max} = \frac{1}{\sigma_b} [E_0 + E_1 - 2\sqrt{E_0 E_1} \rho_{0,1}] \tag{2.30}$$

où $\rho_{0,1}$ est l'intercorrélation normalisée entre les signaux $s_0(n)$ et $s_1(n)$. Lorsque cette intercorrélation est égale à $-1$, les signaux sont les plus dissemblables possibles et on obtient

$$\nu^2_{max} = \frac{1}{\sigma_b} [E_0 + E_1 + 2\sqrt{E_0 E_1}] = \frac{1}{\sigma_b} [E_0 + E_1]^2 \tag{2.31}$$

La probabilité de faire une erreur est alors égale à

$$P_e = Q(\frac{\nu_{max}}{2}) = Q(\frac{\sqrt{E_0} + \sqrt{E_1}}{2\sigma_b}) \tag{2.32}$$

Le récepteur optimale précédent peut être modifié de la façon suivante. Comme

$$v(n) = \sum_k s_1(N-1-k)y(n-k) - \sum_k s_0(N-1-k)y(n-k) \tag{2.33}$$

évalué à l'instant $N-1$, donne

$$v(N-1) = \sum_k s_1(N-1-k)y(N-1-k) - \sum_k s_0(N-1-k)y(N-1-k) \tag{2.34}$$

$$v(N-1) = \sum_k s_1(k)y(k) - \sum_k s_0(k)y(k) \tag{2.35}$$

le récepteur optimale consiste donc à projeter le vecteur d'observation $\underline{y}$ sur les deux vecteurs $\underline{s_0}$ et $\underline{s_1}$ et à choisir celui qui entraîne la plus grande corrélation.

### 2.5.2 Détection dans l'espace de signal

Dans le cas général, on dispose de $M$ signaux $\underline{s_i} = [s_i(0) \dots s_i(N-1)]^t$ associés à $M$ messages distincts. Dans le cas d'une modulation QPSK, $M = 4$. Comme on l'a déjà vu, ces signaux peuvent s'écrire comme combinaison linéaire des éléments d'une base orthonormée de la façon suivante

$$s_i(n) = \sum_{j=0}^{M'-1} a_{ij}\phi_j(n) \tag{2.36}$$

avec $M' \leq M$ et $\{\phi_{ij}\}$ une base orthonormée vérifiant

$$< \underline{\phi_i}, \underline{\phi_j} >= \sum_{n=0}^{N-1} \phi_i(n)\phi_j(n) = \delta(i-j) \tag{2.37}$$

Les coefficients $\{a_{ij}\}$ sont les coordonnées du vecteur $s_i$ sur la base $\phi_j$. Connaissant les vecteurs du dictionnaire, le récepteur peut construire la base orthonormée à l'aide d'une procédure d'orthogonalisation de Gram Schmidt.

Le vecteur observé $\underline{y} = [y(0) \dots y(N-1)]^t$ est un des vecteurs $\underline{s_i}$ entaché du vecteur bruit $\underline{b} = [b(0) \dots b(N-1)]^t$. Appelons $\underline{z} = [z_0 \dots z_{M'-1}]$ la projection de $\underline{y}$ sur les vecteurs de la base. Dans une transmission sans bruit, le vecteur $\underline{z}$ est égal à un des vecteurs $\underline{a_i}$. La connaissance de $\underline{z}$ est suffisante pour en déduire sans erreur le message transmis. Appelons $\underline{c} = [c_0 \dots c_{M'-1}]^t$ la projection de $\underline{b}$ sur les vecteurs de la base. C'est un vecteur aléatoire de moyenne nulle puisque

$$E\{c_i\} = \sum_{n=0}^{N-1} E\{b(n)\}\phi_i(n) = 0 \tag{2.38}$$

et de la matrice de covariance $\sigma_B^2 I$ puisque

$$E\{c_i c_j\} = \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} E\{b(n)b(m)\}\phi_i(n)\phi_j(n) = \sigma_B^2\delta(i-j) \tag{2.39}$$

Le vecteur $\underline{z}$ est donc un vecteur aléatoire de moyenne $\underline{a}_i$ et de matrice de covariance $\sigma_B^2 I$. On recherche quelle est la meilleure décision à prendre connaissant $\underline{z}$ et minimisant la probabilité d'erreur moyenne. Il s'agit de diviser l'espace d'observation engendré par les fonctions de base de façon à décider "le i-ème message a été émis" (hypothèses $H_i$) lorsque le vecteur d'observation tombe dans la partition $D_i$. La règle de décision maximisant la vraisemblance consiste à choisir l'hypothèse $H_i$ qui maximise $p_Z(\underline{z}/H_i)$.

Comme

$$p_Z(\underline{z}|H_i) = \prod_{j=0}^{M'-1}\frac{1}{\sqrt{2\pi\sigma_B^2}}exp(-\frac{(z_j-a_{ij})^2}{2\sigma_B^2}) \tag{2.40}$$

$$p_Z(\underline{z}|H_i) = \frac{1}{(2\pi\sigma_B^2)^{M'/2}}exp(-\frac{\|\underline{z}-\underline{a}_i\|^2}{2\sigma_B^2}) \tag{2.41}$$

La maximisation de cette expression revient à minimiser $\|\underline{z}-\underline{a}_i\|^2$. La meilleur règle de décision consiste donc à chercher parmi $[\underline{a}_0 \ldots \underline{a}_{M'-1}]$ celui qui est plus proche de $\underline{z}$. Le processus de calcul du symbole plus proche est analogue à calculer les corrélations entre le vecteur observé $\underline{y}$ et les différents signaux $\underline{s}_i$ et choisir comme hypothèse celle qui nous donne la corrélation la plus grande. Dans la figure 2.8, on peut voir la procédure de calcul du symbole estimé au moyen d'un banc de corrélateurs.

Dans le cas du système utilisé, le nombre $M$ de symboles du dictionnaire est égal à 4. Les 4 signaux du dictionnaire peuvent être représentés en fonction de deux fonctions $\phi_1, \phi_2$ d'une base orthonormée obtenue par la méthode d'orthogonalisation de Gram Schmidt appliquée au dictionnaire. A partir de $\hat{s}(n)$, il faut déterminer le symbole qui a été envoyé à partir du signal reçu. Pour arriver à cela, le système effectue un processus de corrélation avec les différents signaux du dictionnaire, en utilisant un banc de corrélateurs. Ce processus de corrélation est équivalent à effectuer un processus d'étalement inverse suivi d'un processus de démodulation QPSK, représenté par la figure 2.11

Au récepteur, on prend le même dictionnaire $S$ utilisé à l'émission et donc la même constellation de signaux. Pour décider quel symbole a été émis, on divise l'espace de signaux en 4 régions :

1. L'ensemble des points plus proches du symbole associé au vecteur $s_1(n)$
2. L'ensemble des points plus proches du symbole associé au vecteur $s_2(n)$
3. L'ensemble des points plus proches du symbole associé au vecteur $s_3(n)$

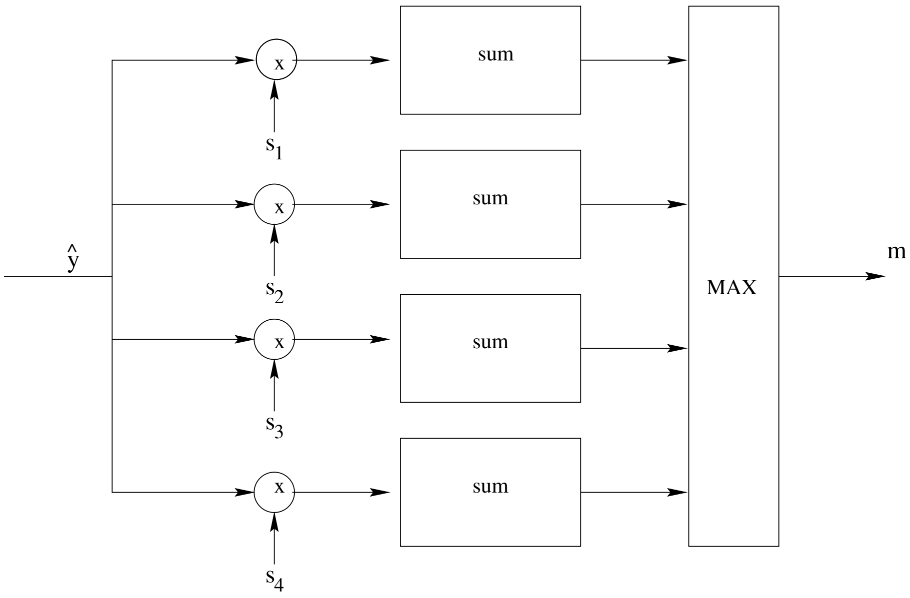


FIG. 2.8 – *Détection : banc de corrélateur*

4. L'ensemble des points plus proches du symbole associé au vecteur $s_4(n)$

On trouve que ces régions de décision sont les quadrants dont les vertices coïncident avec l'origine. Le signal reçu $\hat{s}(n)$ a deux éléments, $s_1$ et $s_2$, résultat de la projection du vecteur $\hat{s}(n)$ sur l'espace de signaux $\phi_i$. Il suffit de projeter le vecteur d'observation $\hat{s}$ sur les vecteurs du dictionnaire et de choisir celui qui entraîne la plus grande corrélation. La figure 2.9 représente la constellation de signaux obtenue pour cette modulation DS/QPSK. On peut observer quatre points correspondant aux quatre symboles du dictionnaire émis. Les nuages de points correspondent aux symboles reçus. On peut voir que les symboles reçus tombent presque toujours dans la région correspondant au symbole émis, la probabilité d'erreur sera donc très faible.

La probabilité d'erreur (BER) est donnée par :

$$P_e = Q(\sqrt{\frac{2E_b}{\sigma_B^2}}) \tag{2.42}$$

Dans la figure 2.10, on voit cette dépendance entre la probabilité d'erreur et le rapport signal bruit. Les conditions de simulation sont les suivantes :

- Quand on parle de signal d'information, il s'agit du signal de tatouage $t(n)$
- Le bruit est représenté par le signal de musique $x(n)$ vu comme un bruit blanc additif de moyenne nulle et densité spectrale de puissance $\frac{N_0}{2}$.
- Il n'y a pas d'étalement de spectre effectué sur le signal de tatouage ($N_c = 1$).

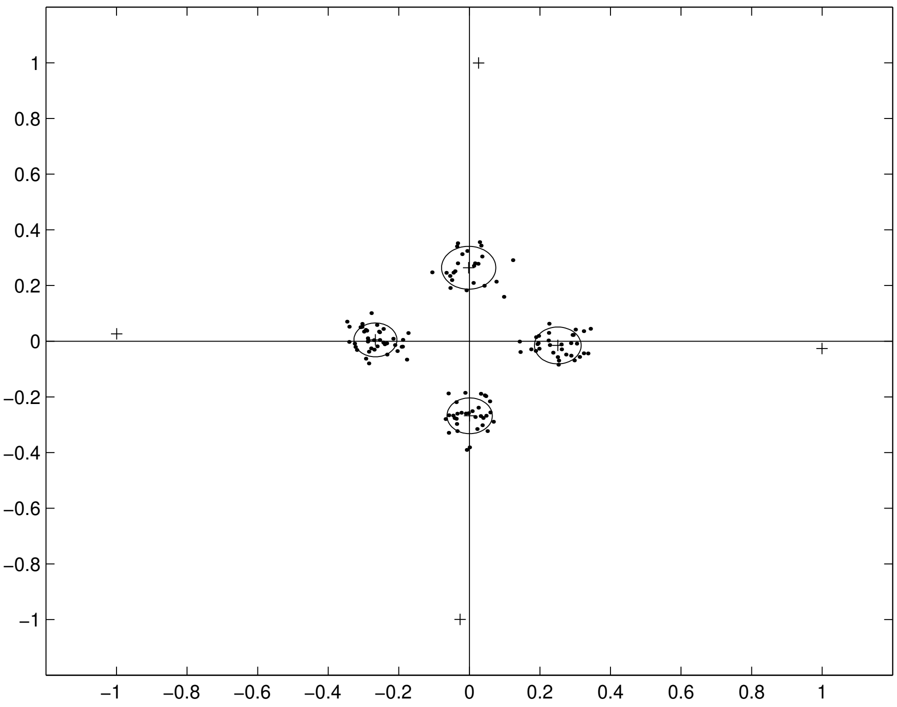


FIG. 2.9 – *Représentation des symboles émis et reçus dans l'espace de signaux*

### 2.5.3 Etalement inverse en réception

Dans la détection, il faut effectuer une démodulation et un processus d'étalement inverse au réalisé en transmission. Le processus de détection est composé, donc, des deux taches représentées dans la figure 2.11. L'opération d'étalement inverse en réception est effectuée sur le signal reçu, qui est la somme du signal envoyé plus le bruit du canal $\hat{y}(n) = y(n) + q(n)$. Le spectre du bruit est donc aussi étalé. Comme l'énergie du signal reste constante au cours de l'étalement, l'amplitude de la densité de puissance du bruit additif $S_b(f)$ devient plus petite que dans le cas sans d'étalement. Le processus de filtrage en réception élimine le bruit hors de la bande étroite du signal. La puissance du bruit après le processus étalement-filtrage a donc été divisée par $N_c$. En terme de probabilité d'erreur, cela signifie une meilleur performance.

On peut montrer que la probabilité d'erreur dans le cas où on effectue un étalement de spectre peut s'exprimer de la façon suivante:

$$P_e = Q(\sqrt{\frac{2E_b}{JT_c}}) \tag{2.43}$$

où $E_b$ représente l'énergie bit, $T_c = \frac{T_b}{2\ N_c}$ la durée de chaque bit de la séquence PN (qui est inversement proportionnel à la longueur de la séquence $N_c$) et J la puissance moyenne

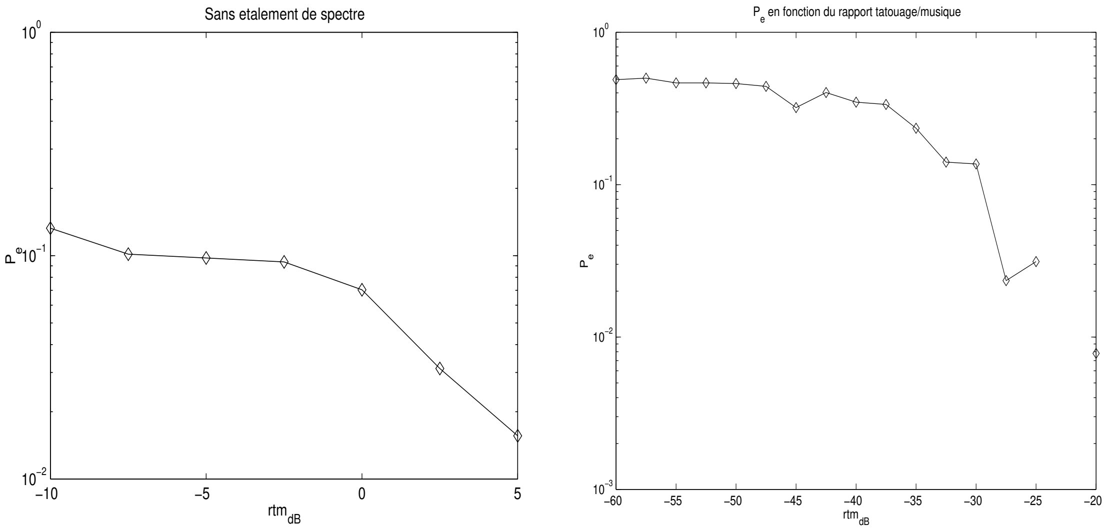


FIG. 2.10 – *Taux d'erreur en fonction de la relation signal bruit (tatouage/musique). A gauche : sans étalement de spectre. A droite : avec étalement de spectre.*

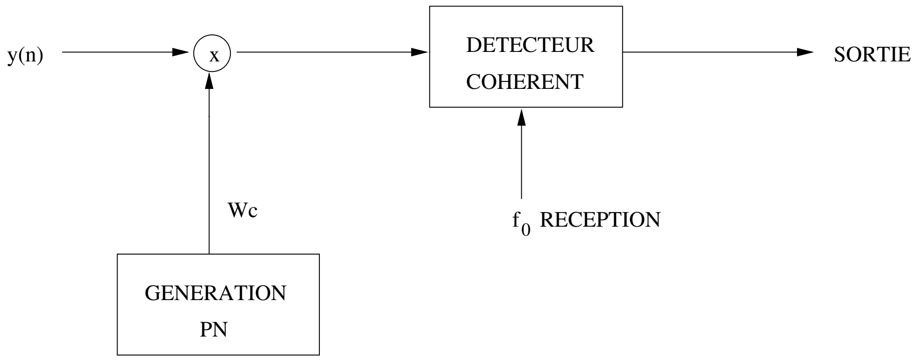


FIG. 2.11 – *Processus de détection*

du signal d'interférence. On obtient

$$P_e = Q(\sqrt{\frac{2E_b}{JT_c}}) = Q(\sqrt{\frac{4E_bN_c}{JT_b}}) \tag{2.44}$$

L'interférence peut se voir comme un bruit blanc de densité spectrale de puissance $\frac{\sigma_B^2}{2}$, où

$$\frac{\sigma_B^2}{2} = \frac{JT_c}{2} \tag{2.45}$$

En prenant l'expression de la probabilité d'erreur, on observe sur le tracé de droite de la figure 2.10 que la probabilité d'erreur diminue quand on augmente $N_c$. On a le même taux d'erreur pour un rapport signal sur bruit beaucoup plus faible.

Pour que la détection soit correcte, l'émetteur et le récepteur doivent être synchronisés et utiliser toujours la même séquence pseudoaléatoire. Sinon, le taux d'erreur devient très élevé.

# Chapitre 3

# Paramètres déterminant la modulation DS/QPSK

A l'émetteur, pour construire le signal de tatouage, on effectue un étalement de spectre suivi d'une modulation QPSK (cf figure 2.2). Au récepteur, il faut effectuer le processus inverse pour retrouver l'information cachée (cf figure 2.11). Pour construire le dictionnaire associé et donc l'espace de signaux, il faut connaître les paramètres suivants :

- la taille $N_c$ de la séquence pseudoaléatoire utilisée dans l'étalement de spectre ainsi que les valeurs $\{W_c\}$ de cette séquence,
- la fréquence de la porteuse utilisée dans la modulation BPSK.

En fonction de ces deux paramètres, on obtient le dictionnaire (constellation de signaux) $S(W_c, f_0)$.

## 3.1 Changement de la séquence pseudoaléatoire

C'est possible de changer de séquence pseudoaléatoire au cours de la transmission de symboles. Si la détection ne s'effectue pas en utilisant la même séquence pseudoaléatoire pour l'étalement de spectre inverse, on aura un taux d'erreur important. La copie locale de la séquence pseudoaléatoire du récepteur doit aussi être synchronisée avec la séquence utilisée pour coder le tatouage.

D'après quelques expériences réalisées, si le récepteur n'utilise pas la même séquence pseudoaléatoire que le transmetteur ($S(W_{cTx}, f_0) \neq S(W_{cRx}, f_0)$), le taux d'erreur monte à $P_e \approx 0.6$. La détection ne se fait donc pas correctement.

## 3.2 Changement de la fréquence de la porteuse

Nous avons déjà vu que la fréquence de la porteuse utilisée pour moduler le signal de tatouage doit être comprise entre les limites données par l'équation 2.2 afin d'éviter le recouvrement de spectre entre les deux lobes principaux. Le récepteur doit décoder en utilisant la même porteuse que celle utilisée en transmission pour avoir des bonnes performances.

Dans la figure 3.1, on remarque le comportement suivant. Le transmetteur utilise une

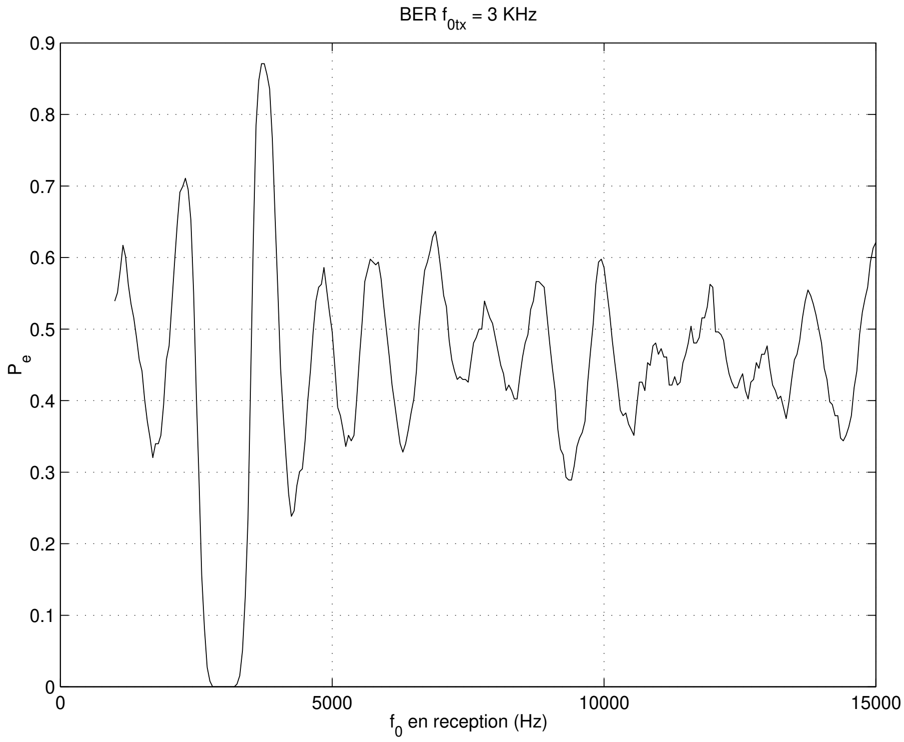


FIG. 3.1 – *Taux d'erreur en fonction de la fréquence utilisée en réception*

fréquence de porteuse fixe. Si le récepteur utilise une fréquence de porteuse différente de celle utilisée pour coder le tatouage, le taux d'erreur augmente considérablement. Dans cet exemple, la fréquence utilisée en transmission est $f_0 = 3$ kHz et celle utilisée en réception varie entre les limites établis auparavant : $f_0 \,\epsilon\, [1, 15]$ kHz avec une fréquence d'échantillonnage de $f_e = 32$ kHz. On a utilisé un étalement de spectre fixe pour toute l'expérience $N_c = 8$. Le taux d'erreur varie entre 0.3 et 0.6 pour toute fréquence différente de celle utilisée en transmission. On voit qu'il y a une zone autour de 3 kHz (porteuse en transmission) où le taux d'erreur est faible. La largeur de cette zone est égale à la largeur de bande du signal de tatouage, c'est à dire la largeur du lobe principal. Dans cet exemple, cette largeur est égale à :

$$\triangle f \approx N_c \frac{2}{T_{symbole}} = 8 \frac{2 * 32000}{2 * 256} = 1 \text{ kHz} \tag{3.1}$$

Dans la figure 3.2, on a représenté cette variation du taux d'erreur pour trois longueurs différentes de la séquence pseudoaléatoire : le cas où il n'y a pas d'étalement ($N_c = 1$) et pour les valeurs $N_c = 8$ et $N_c = 16$. On peut constater que la largeur du lobe principal

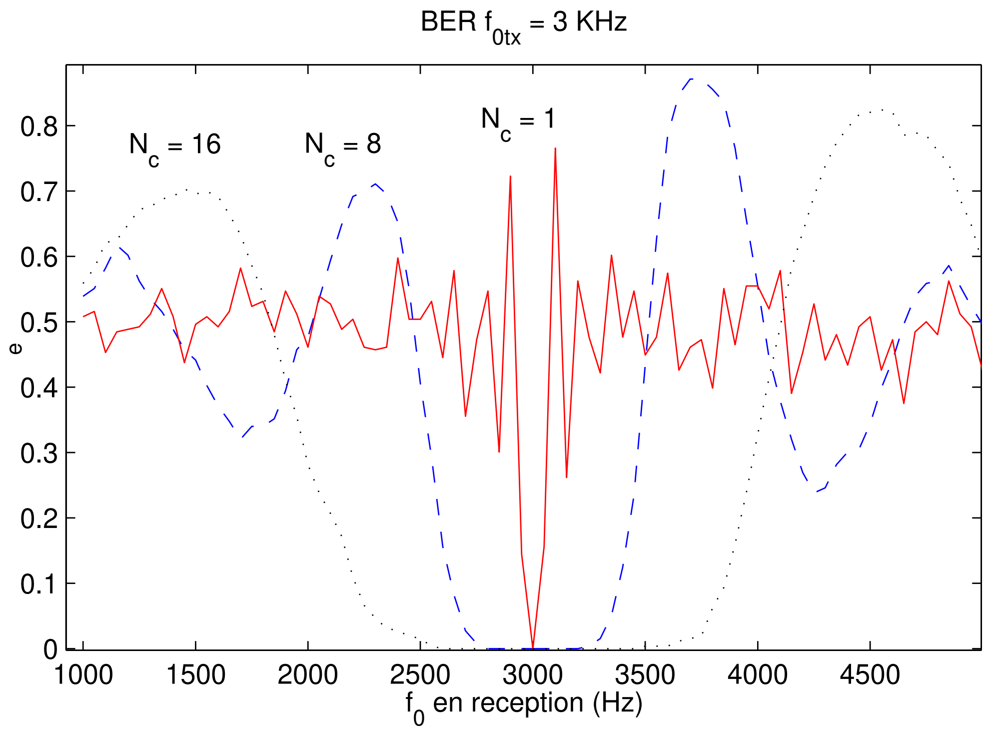


FIG. 3.2 – *Taux d'erreur en fonction de la fréquence utilisée en réception pour différentes longueurs de la séquence pseudoaléatoire :* $N_c = 1, 8, 16$

est proportionnelle à la valeur de $N_c$.

## 3.3 Schéma de tatouage

On a défini deux paramètres qui déterminent le type de modulation employée et qui sont utilisés dans la construction du dictionnaire. On aura donc un dictionnaire différent pour chaque paire de valeurs $(W_c, f_0)$. Le récepteur, pour être capable de décoder le signal de tatouage, doit calculer la corrélation du signal reçu avec les différents signaux $\{s_i(n)\}$, $i = 1, \ldots, M$ appartenant au dictionnaire. Il a besoin donc de connaître le dictionnaire utilisé pour coder les symboles. Si le récepteur utilise un dictionnaire différent de celui utilisé en transmission, le taux d'erreur devient important, il faut donc avoir une très bonne synchronisation des dictionnaires en transmission et en réception. Ce besoin de synchronisation peut être utilisé pour cacher de l'information.

En effet, supposons que nous changions les paramètres de la modulation au cours du temps selon deux lois d'enchaînement comme le symbolise le schéma de la figure 3.3.

Soient $\{f_{0i}\}$, $i = 1, \ldots, I$ l'ensemble des différentes porteuses utilisées et $\{W_{cj}\}$, $j =$

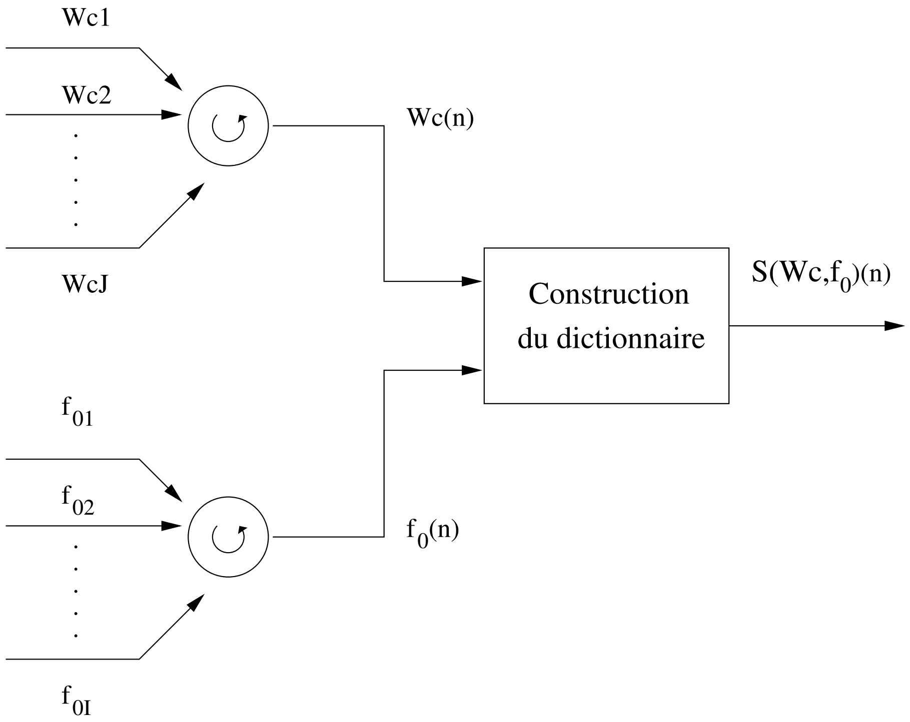


FIG. 3.3 – *Schéma de génération du dictionnaire*

$1, \ldots, J$ l'ensemble des différentes séquences pseudoaléatoires possibles, de longueur $\{N_{cj}\}$, $j = 1, \ldots, J$. En employant ces différentes valeurs, on trouve l'ensemble des dictionnaires possibles $\{S(W_{ci},\ f_{0j})\}$.

En principe, on change de dictionnaire à chaque symbole émis.

Le récepteur doit connaître cet enchaînement de paramètres pour être capable de décoder correctement les données. Néanmoins, si le récepteur connaît les valeurs possibles $\{N_{cj}\}$, il peut construire toutes les séquences possibles $\{W_{cj}\}$, tous les dictionnaires associés à ces séquences et il peut calculer les fonctions de corrélation. En prenant le symbole qui entraîne la plus grande corrélation, il peut arriver à trouver le symbole émis. Pour cela, il faut quand même calculer $N = 2^{N_{ci}}$ dictionnaires différents pour chaque symbole, ce qui devient trop coûteux au niveau calcul pour des valeurs élevées de $N_{ci}$.

Ensuite le récepteur, en connaissant la séquence pseudoaléatoire utilisée, peut calculer un périodogramme moyenné du signal après l'avoir multiplié par la séquence (opération d'étalement inverse ou *despreading*). Les pics de ce périodogramme nous indiquent les différentes fréquences porteuses utilisées. L'enchaînement de ces fréquences reste en principe inconnu. Mais si on construit les dictionnaires correspondants à toutes les fréquences possibles, on calcule toutes les corrélations et en en prenant le maximum, on peut trouver le symbole envoyé avec une probabilité d'erreur très faible.

# Chapitre 4

# Méthodes de Synchronisation

Jusqu'à maintenant, nous avions considéré une synchronisation parfaite entre le récepteur et le transmetteur de l'information de tatouage mais ce n'est pas forcément le cas. Dans ce chapitre, on se pose le problème de la synchronisation. Pour effectuer correctement la détection, le récepteur doit connaître l'instant où la modulation peut changer d'état. C'est-à-dire, il doit connaître les débuts et les fins des symboles afin de commencer la détection. L'estimation de ces instants s'appelle *synchronisation de symbole*. De la même manière, les systèmes utilisant la technique d'étalement de spectre ont besoin que la séquence pseudoaléatoire générée localement au niveau du récepteur soit synchrone à celle utilisée pour étaler le signal en transmission. Une *synchronisation de séquence pseudoaléatoire* est donc aussi nécessaire pour une synchronisation parfaite.

Dans ce chapitre, nous allons nous poser les problèmes de la synchronisation de symbole qui consiste à rechercher les mêmes limites de fenêtres pour la détection et pour la transmission. Le pirate peut enlever des échantillons du signal de façon à perdre la référence pour le fenêtrage et donc la synchronisation de symbole. On propose une méthode de synchronisation par des méthodes de corrélation.

## 4.1 Opérations de piratage

Le pirate peut effectuer des modifications sur le signal reçu (musique + tatouage) qui font que le récepteur perd le synchronisme de symbole. Il peut enlever des échantillons au hasard de façon que le décodeur n'ait plus sa référence de fenêtrage (échantillons de début et de fin de fenêtre).

Les modifications que le pirate effectue sur le signal ne doivent pas être audibles. D'après quelques expériences réalisées, le pirate peut enlever en moyenne au maximum un échantillon tous les 2500 échantillons environ sans que cela soit perceptible. Néanmoins, dans les morceaux du signal où il y a des silences, le pirate peut se permettre d'enlever beaucoup plus d'échantillons sans que l'auditeur s'en aperçoive. Inversement, le pirate peut rajouter des échantillons par interpolation. Des expériences ont montré que les ordres de grandeur sont approximativement identiques (1 échantillon tous les 2500), même si l'addition d'échantillons par interpolation est moins audible.

Chaque fois que le pirate effectue ce type d'opérations d'ajout ou de retrait d'échantillons au signal tatoué, il introduit un délai entre ce signal et la référence du récepteur de l'information. Ce délai est égal au nombre d'échantillons enlevés (ou rajoutés).

## 4.2 Synchronisation initiale

Au début du processus de détection des données cachées, il faut établir une synchronisation de symbole pour pouvoir les décoder correctement. L'une des techniques les plus efficaces utilisés en communications numériques consiste à transmettre une séquence de données connue par le transmetteur et le récepteur. Cette séquence doit être suffisamment courte pour permettre de chercher, dans un temps raisonnable, le synchronisme dans toutes les positions possibles de symbole et suffisamment longue pour pouvoir satisfaire les exigences des fonctions de corrélation. Le temps de synchronisation est égal à la durée de cette séquence de données, appelée *séquence d'apprentissage*.

Le but du processus de synchronisation initiale est d'estimer le délai qui a été introduit par le pirate, c'est-à-dire le déplacement relatif entre les positions de symboles dans le signal de tatouage et les limites des fenêtres d'analyse utilisées en réception. Dans le cas du système de tatouage utilisé, on fait une analyse par fenêtres de longueur $N = 512$ échantillons en transmettant 1 symbole parmi 4 par fenêtre ($R_b = 125$ bps). Le problème de la synchronisation consiste donc à calculer le déplacement du début de symbole par rapport à la fenêtre utilisée. Pour calculer le délai introduit, on se sert d'une fenêtre d'analyse glissante se déplacant un nombre $k$ d'échantillons en avant ou en arrière avec $k \in [-K, K]$ comme le montre le schéma de la figure 4.1.

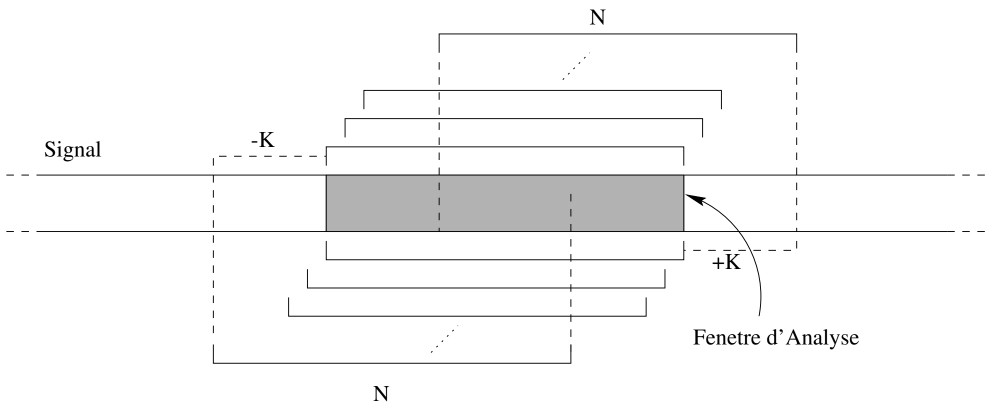


FIG. 4.1 – *Schéma d'utilisation d'une fenêtre glissante avec déplacement $k \in [-K, K]$*

La séquence d'apprentissage est composée d'une suite de signaux $s_i(n)$, appartenant au dictionnaire $S$ utilisé. Pour chaque valeur $k$ du déplacement de la fenêtre glissante, la corrélation entre le signal reçu $\hat{s}(n)$, $n\epsilon[k, \ldots, k+N-1]$ et le signal du dictionnaire associé au symbole, $s_i(n), n\epsilon[0, \ldots, N-1]$, prend une valeur différente.

**Caractéristiques de la fonction de corrélation**

Pour étudier les caractéristiques de la fonction de corrélation du signal tatoué, on utilise la fenêtre glissante décrite précédemment dans un signal qui n'a pas été modifié et qui est donc parfaitement synchrone au récepteur. Le tracé de la figure 4.2 montre comment varie la fonction de corrélation en fonction de $k$ pour une fenêtre d'analyse donnée. La forme de

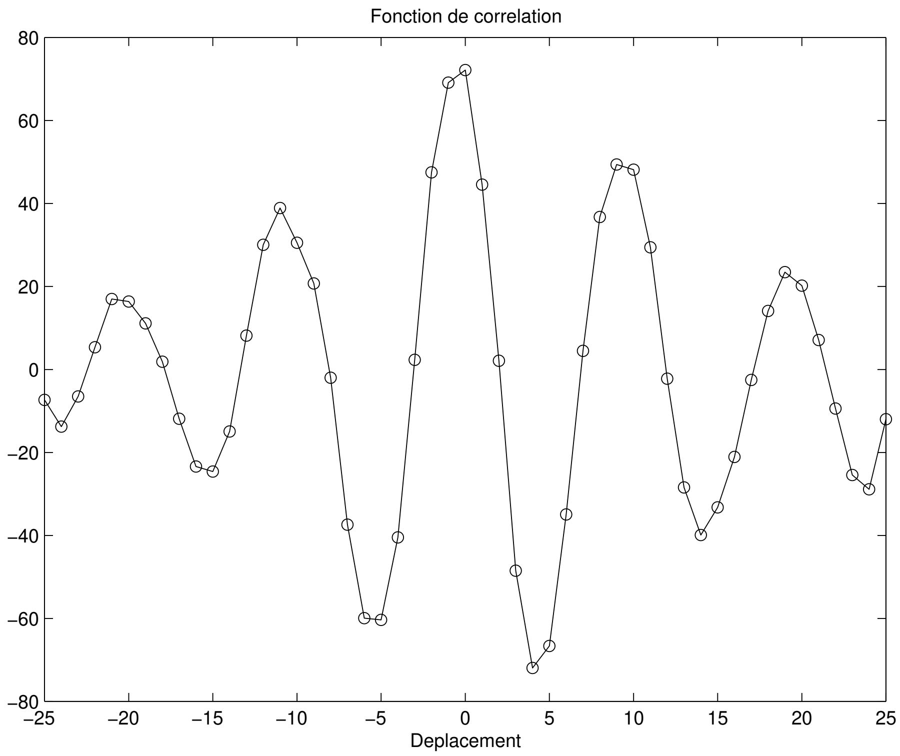


FIG. 4.2 – *Fonction de corrélation en fonction du déplacement de la fenêtre glissante*

cette fonction de corrélation est caractéristique de la modulation utilisée. Dans le système de tatouage étudié, la modulation employée est une modulation de phase QPSK d'une porteuse de fréquence $f_0$. La corrélation maximale se trouve pour $k = 0$ et s'appelle *pic de corrélation*. Le nombre d'échantillons séparant deux pics consécutifs de la fonction de corrélation est fixée par la fréquence de la porteuse et la fréquence d'échantillonnage par la relation suivante :

$$N_0 = \frac{f_e}{f_0} \tag{4.1}$$

La condition de Nyquist impose que $N_0 > 2$. Avec une fréquence d'échantillonnage de 32 kHz, $N_0$ diminue avec la fréquence de la porteuse de la façon suivante :

| $f_0$ (kHz) | $N_0$ |
|---|---|
| 1 | 32 |
| 3 | 10.7 |
| 5 | 6.4 |
| 7 | 4.6 |
| 9 | 3.6 |
| 11 | 2.9 |
| 13 | 2.5 |
| 15 | 2.1 |

On a donc intérêt à utiliser une faible fréquence de porteuse pour faciliter la synchronisation. On peut remarquer d'autre part que la courbe qui enveloppe les maxima est déterminée par l'étalement de spectre utilisé.

Le tracé de la figure 4.3 montre comme varie la position du pic de corrélation au cours des différentes fenêtres d'analyse. Nous pouvons vérifier que, même si le pic n'est pas toujours centré, il se trouve en moyenne pour une valeur nulle du déplacement ($k \approx 0$), comme montre la figure 4.6.

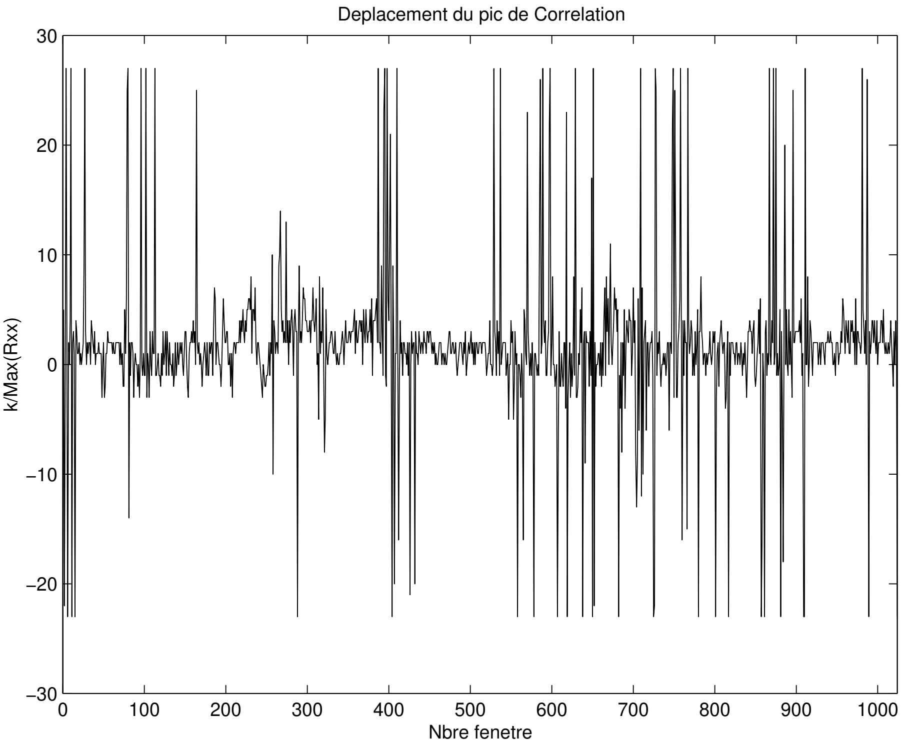


FIG. 4.3 – *Position du pic de corrélation pour les différents symboles de la séquence d'apprentissage*

Dans cette expérience, une séquence d'apprentissage connue de 1024 symboles a été transmise. La valeur de ce maximum n'est pas non plus constante, mais l'information donnée par cette variation du pic de corrélation n'est pas utilisée pour la synchronisation.

### Calcul du délai introduit

Jusqu'à maintenant, le récepteur restait synchrone au signal et le délai était donc nul. Maintenant, on traite le cas d'une perte de synchronisme, mais on se limite, pour l'instant, au cas (peu réaliste) où le pirate ne supprimerait que quelques échantillons au début di fichier. Le récepteur essaie d'estimer le déplacement entre sa fenêtre d'analyse et le début de symbole. Pour réaliser ce calcul, il se sert des caractéristiques de la fonction de corrélation.

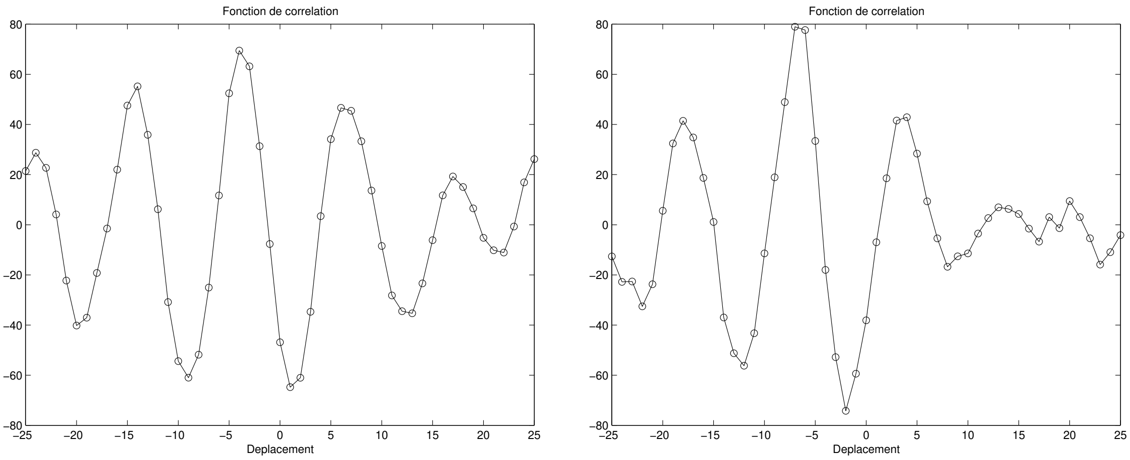


FIG. 4.4 – *Fonctions de corrélations avec un délai de 3 (gauche) et 6 (droite) échantillons*

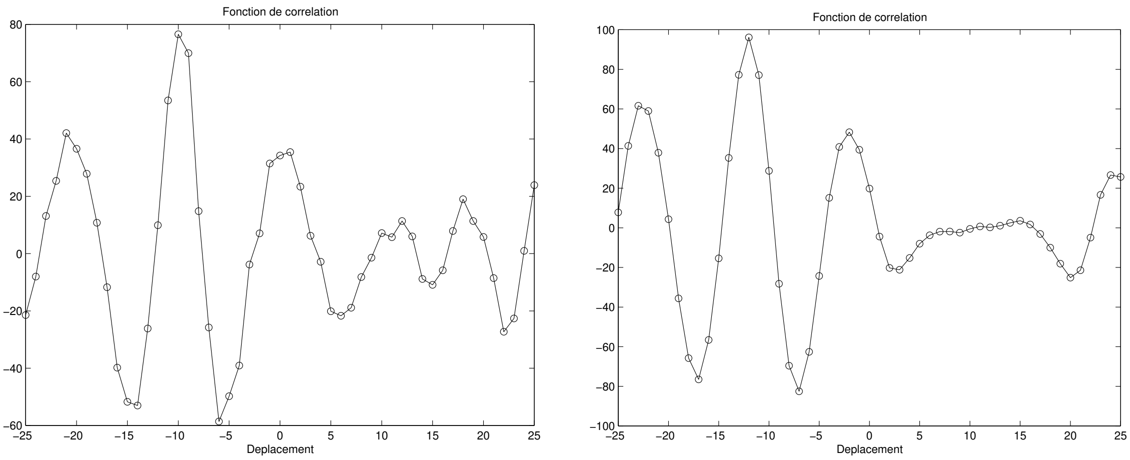


FIG. 4.5 – *Fonctions de corrélations avec un délai de 9 (gauche) et 12 (droite) échantillons*

L'utilisation d'une fenêtre glissante et d'une séquence d'apprentissage, nous amène aux résultats suivants : les figures 4.4 et 4.5 montrent que le pic de corrélation se déplace d'une valeur environ égale au délai entre les fenêtres de transmission et les fenêtres de réception.

Le processus de synchronisation consiste donc à calculer le déplacement du pic de corrélation ($\hat{k}$) puis à déplacer la fenêtre d'analyse d'un nombre $\hat{k}$ d'échantillons pour se synchroniser. Néanmoins, il faut un certain temps pour calculer la moyenne du déplacement. Ce temps de calcul détermine donc le temps d'acquisition du synchronisme.

Le tracé de la figure 4.3 montrait comment le déplacement du pic de corrélation variait autour de la valeur du délai (0 dans ce cas). Les deux tracés de la figure 4.6 montrent l'évolution du déplacement moyen du pic de corrélation. Il tend vers la valeur du délai introduit (0 à gauche, 3 à droite).

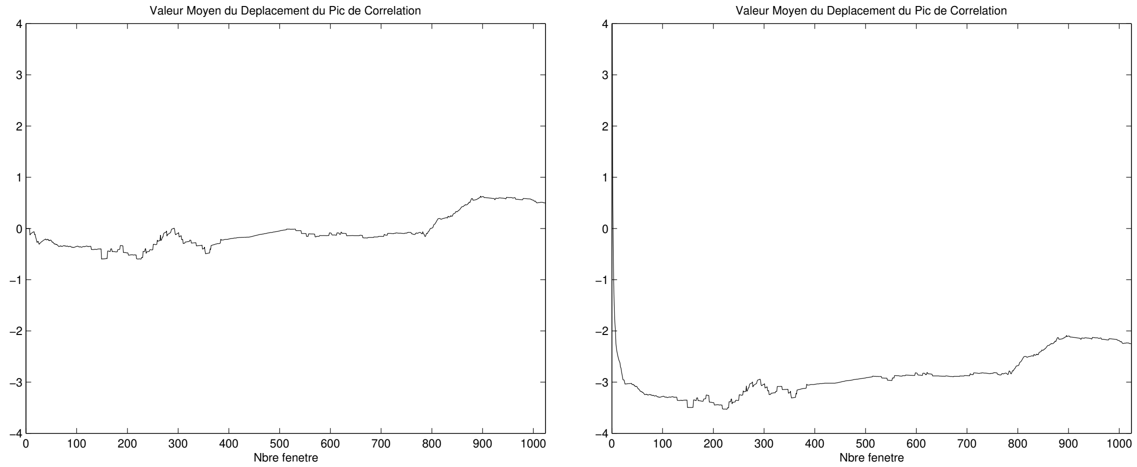


FIG. 4.6 – *Evolution de la moyenne du délai au cours de la séquence d'apprentissage*

Supposons qu'à partir de la fenêtre $i = I$, la valeur moyenne se stabilise à la valeur du délai. Le temps d'acquisition du synchronisme est donc égal à $I$ fenêtres, pour une fréquence d'échantillonnage de $f_e$ et une porteuse à $f_0$.

Ce temps d'acquisition varie en fonction de la fréquence de la porteuse et du retard introduit, dû aux propriétés des fonctions de corrélation du signal.

Une séquence d'apprentissage bien choisie est une bonne solution pour presque tous les problèmes de synchronisation. Le temps d'acquisition est fixé par la durée de cette séquence et doit s'ajouter au début de chaque transmission. Les séquences de synchronisation (appelées aussi préambules) ont une durée comprise entre quelques centaines et quelques milliers de bits dans les systèmes de transmission typiques, en fonction des besoins du système. Si la séquence de synchronisation est courte, ce qui permet une synchronisation rapide, le système est plus vulnérable aux fausses corrélations et à la possible reproduction à l'aide d'une interférence. A l'exception de ce problème de vulnérabilité, cette technique de synchronisation est présenté l'avantage d'être simple à implémenter.

## 4.3 Poursuite

Dans le cas précédent, le pirate avait introduit un certain délai au signal tatoué. Le déplacement du pic de corrélation était fixe et le récepteur prenait un certain temps de synchronisation pour le calculer. Supposons maintenant qu'il ne s'agit pas d'un retard localisé mais que plusieurs délais ont été introduits à différents points du signal en respectant les limites des modifications inaudibles.

L'objectif est de pouvoir se synchroniser au cours du processus de détection en utilisant l'information donnée par le déplacement des pics de corrélation et ainsi éviter l'utilisation d'une séquence d'apprentissage connue. Si on l'utilise, le pic de corrélation va se déplacer en moyenne en fonction des retards introduits. Dans la figure 4.7, on voit comment se déplace le pic de corrélation quand le pirate introduit un délai de 5 échantillons chaque 256 fenêtres, et dans la figure 4.8 on voit le comportement quand le pirate enlève 8 échantillons chaque 241 fenêtres.

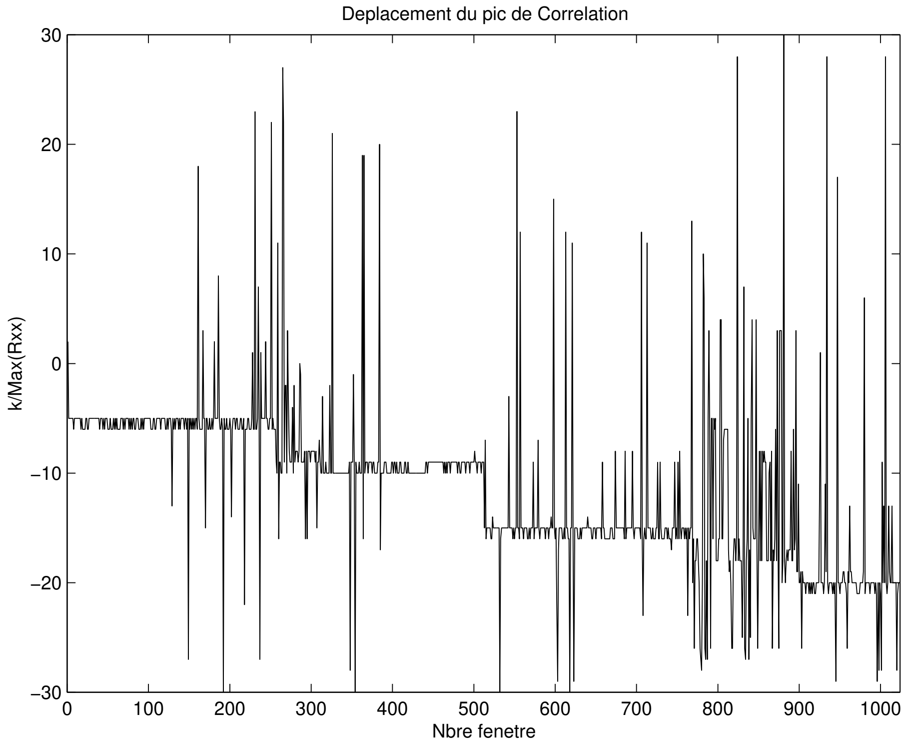


FIG. 4.7 – *Index de corrélation maximale pour une suppression de 5 ech/256 fenêtres*

Supposons maintenant que la séquence de données transmise n'est pas connue par le récepteur. C'est ceci qui distingue la méthode de poursuite de la méthode de synchronisation initiale. Le récepteur devra donc calculer, dans toute la fenêtre glissante, les fonctions de corrélation associées à chaque symbole appartenant au dictionnaire. Dans le cas d'une modulation QPSK, il calcule les fonctions de corrélation associées aux 4 signaux du dic-

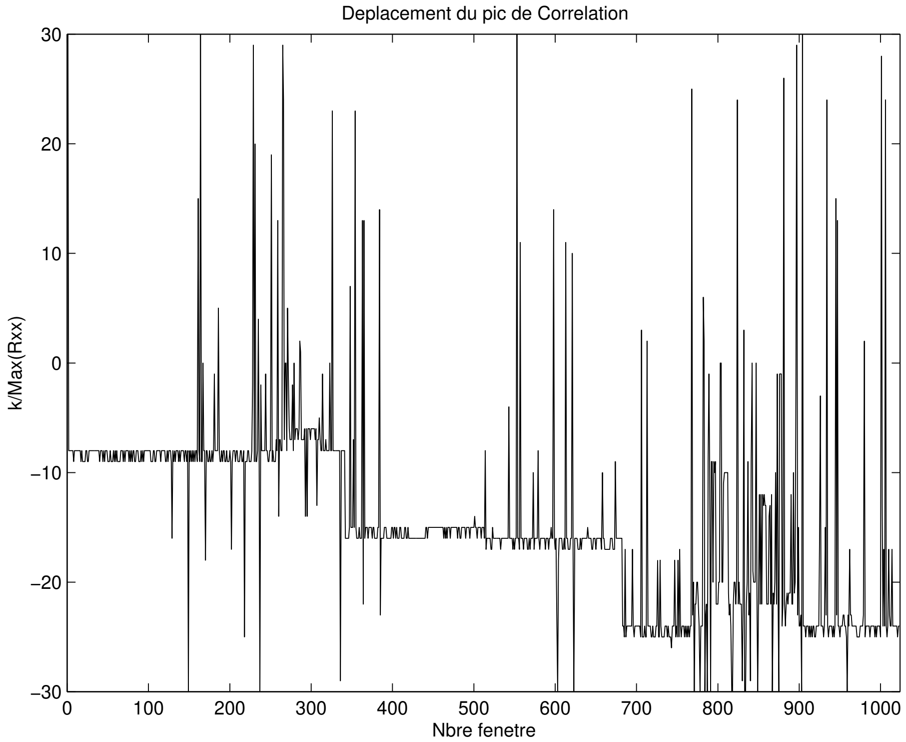


FIG. 4.8 – *Index de corrélation maximale pour une suppression de 8 ech/341 fenêtres*

tionnaire pour les différentes valeurs du déplacement de la fenêtre glissante. Dans la figure 4.9 on trouve la représentation des différentes fonctions de corrélation en fonction du déplacement $k$.

Pour décider quel symbole a été transmis, le récepteur calcule le maximum de ces 4 fonctions de corrélation $\{R_{\hat{s}\hat{s}}(k)\}i\epsilon[1,\ldots,4]\}$ et estime le délai $\hat{k}$ et le symbole $i$ transmis. L'utilisation d'une fenêtre glissante peut générer des erreurs de détection à cause du délai. Un exemple est représenté sur la figure 4.10. Le délai introduit cette fois ci est de $k = -3$ et les fonctions de corrélation sont donc déplacées de 3 échantillons vers la gauche. La fonction de corrélation associée au symbole émis est maximale pour $k = -3$ mais il y a une autre valeur supérieure pour $k = 0$ qui ne correspond pas au symbole tranmis. On aura donc une erreur de détection en considérant toutes les valeurs de la fenêtre glissante.

Comme dans le cas de la synchronisation initiale, le récepteur estime le délai en calculant la valeur moyenne du déplacement du pic de corrélation. La procédure de poursuite serait donc la suivante : chaque $N_t$ symboles, on calcule le déplacement moyen du pic de corrélation dans les derniers $N_t$ symboles : c'est l'estimée du délai introduit dans cette période $\hat{k}$. Si cette moyenne est non nulle, on se déplace de $\hat{k}$ échantillons pour retrouver le synchronisme de symbole. Dans le cas où un délai de quelques échantillons par fenêtre a été introduit, le récepteur essaie de suivre ce délai en calculant cette moyenne et il obtient le résultat montré par la figure 4.11.

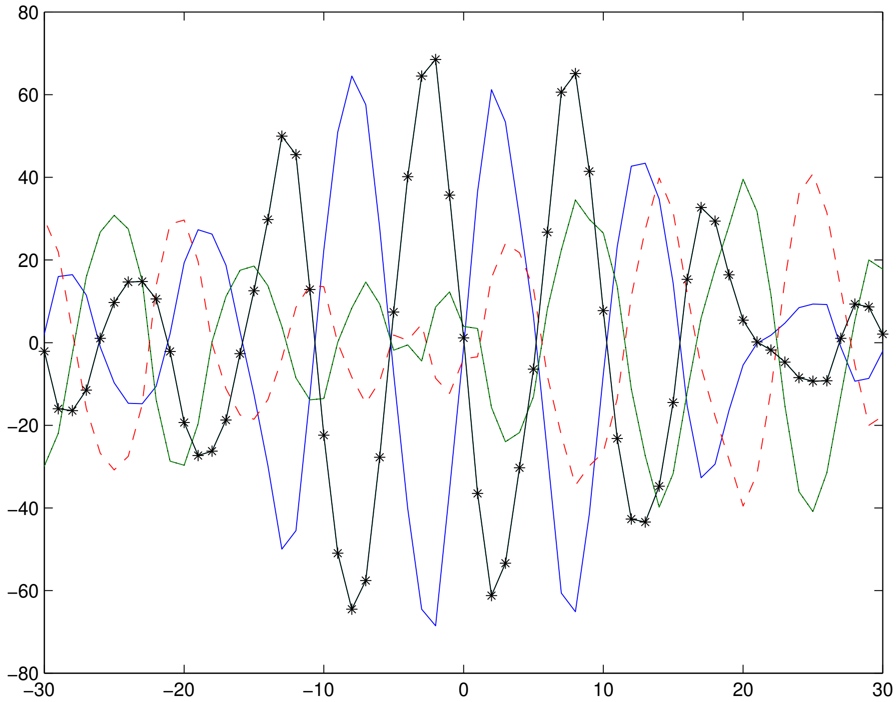


FIG. 4.9 – *Fonction de corrélation pour les 4 symboles du dictionnaire DS/QPSK*

Néanmoins, la séquence de données transmise n'étant pas connue, la probabilité d'erreur dans la détection devient plus grande et le temps de synchronisation augmente considérablement.

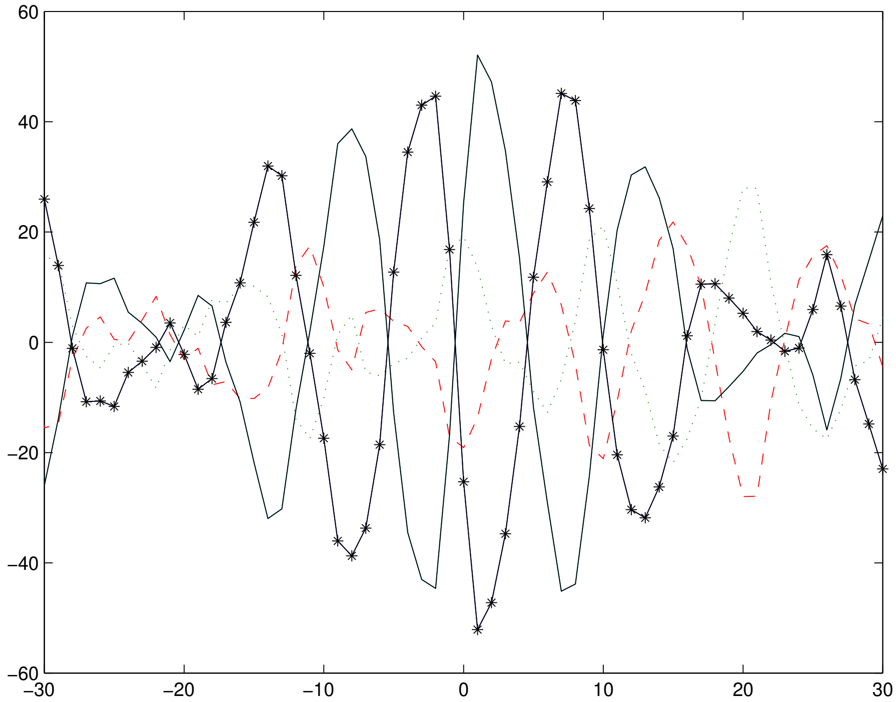


FIG. 4.10 – *Fonction de corrélation pour les 4 symboles du dictionnaire DS/QPSK. Erreur de détection [− ∗ −] symbole transmis [—] symbole décidé*

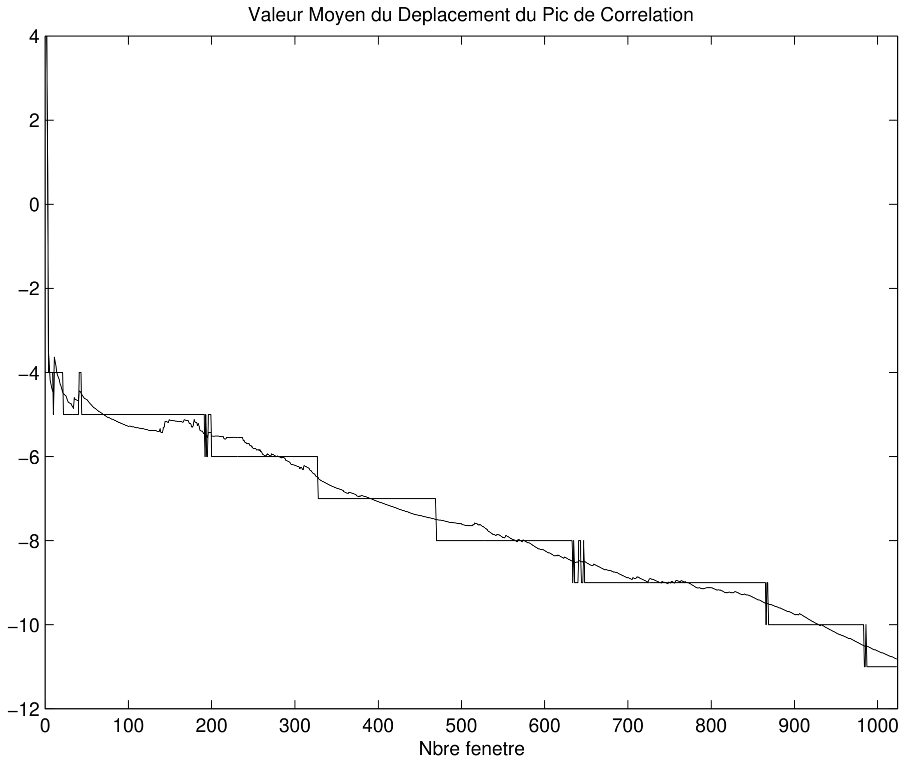


FIG. 4.11 – *Variation de la valeur moyenne du déplacement quand il y a un délai de 5 échantillons chaque 256 fenêtres*

# Chapitre 5

# Conclusion

Le tatouage de sons consiste à cacher de l'information dans le signal audio tel que seules les personnes autorisées puissent y accéder pour la lire ou pour effectuer des modifications. Néanmoins, une personne non autorisée, le pirate, essaiera de retrouver cette information pour l'effacer, la modifier, ou pour empêcher son décodage. C'est pour cette raison qu'une contrainte difficile à respecter dans un système de marquage sonore est la robustesse devant les diverses attaques ou modifications possibles.

En première analyse, nous avons vu comment certains paramètres déterminant la modulation employée (fréquence de la porteuse et étalement de spectre) peuvent être utilisés pour protéger l'information cachée. Si le décodeur ne connaît pas l'enchaînement des paramètres utilisés, il est incapable, en principe, de décoder cette information. Néanmoins, on a vu aussi comment le pirate peut retrouver cet enchaînement en essayant toutes les valeurs possibles des paramètres. Il faudrait donc utiliser en complément d'autres techniques pour la rendre plus robuste à une détection autorisée.

En se mettant toujours à la place du pirate, nous avons modélisé certains attaques inaudibles qu'il peut effectuer sur le signal et qui essaient de faire en sorte que le récepteur ne puisse pas retrouver l'information, car ces attaques lui font perdre sa réference de synchronisme. Dans une deuxième étape, en se mettant cette fois-ci à la place du récepteur autorisé, nous avons trouvé que, par l'analyse des caractéristiques de corrélation du signal, le décodeur peut détecter ces modifications, les quantifier et les contrer d'une certaine façon. Comme résultat, on a defini deux procedures de synchronisation qui permettent au récepteur autorisé de se synchroniser au flux de données reçu. Le système devient donc plus résistant à ce type d'attaques.

Pour concevoir des procédures de tatouage, il faut voir le problème depuis deux points de vue différents. D'une fait la vision du propriétaire du signal, qui essaie de marquer ses signaux, et d'autre fait la vision du pirate, qui tente d'effacer cette information de tatouage. Cette double vision nous permet de trouver des techniques plus résistantes aux attaques.

# Annexe A

# Modèle Psychoacoustique MPEG-1

Les signaux audio sont composés par des signaux de parole de qualité téléphonique, des signaux de parole de bande large et des signaux audio de large bande. La gamme de fréquences pour ce type de signaux audio est de $300 - 3400Hz$ pour des signaux de parole téléphonique, $50 - 7000Hz$ pour des signaux de parole de bande large, et $20 - 20000Hz$ pour des signaux audio de haute qualité et de bande large.

Le système auditif humain peut être modelisé par un ensemble de 26 filtres passe-bande avec des largueurs de bande qui augmentent avec la fréquence. Les 26 bandes s'appellent les bandes critiques. Chaque bande critique est définie autour d'une fréquence centrale dont la largueur de bande du bruit est diminuée jusqu'à ne pas avoir de différence audible avec le ton placé dans la fréquence centrale. Si un ton faible tombe dans la bande critique d'un ton plus fort, le ton faible sera inaudible.

Des modèles de masquage fréquentiel ont été définis pour le codage perceptuel de signaux audio parce qu'il n'est pas nécessaire de coder des informations irrélevantes au niveau perceptif.

Dans cette étude, le modèle psychoacoustique utilisé est le même que celui défini dans la norme MPEG-1 pour une fréquence d'échantillonnage de $32KHz$. Ce modèle peut aussi être utilisé pour des fréquences d'échantillonnage de $44.1kHz$ et $48KHz$.

Le modèle psychoacoustique utilisée en MPEG-1 s'implémente selon la procédure suivante :

- Première phase :

  Soit le signal $s(n)$ celui qui va être analysé. Il faut le diviser en fenêtres de $N = 512$ échantillons du signal $s(n), n = 1, \ldots, N$ Chaque segment (de durée $16ms$) est multiplié échantillon par échantillon par des coefficients appartenant à une fenêtre de Hann, $h(n)$ :

$$h(n) = \frac{\sqrt{\frac{8}{3}}}{2} * [1 - cos(2\pi \frac{n}{N})]$$

Le spectre de puissance du signal $s(n)$ peut se calculer de la façon suivante :

$$S(k) = 10 * log_{10}[\frac{1}{N}||\sum_{n=0}^{N-1} s(n)h(n)exp(-2j\pi\frac{nk}{N}||^2]$$

Le maximum est normalisé par rapport à un son de référence de niveau de pression $96dB$.

- Deuxième phase : il s'agit de l'identification des composantes tonales (partie sinusoïdale) et des composantes pas tonales (bruit). Etant donné que ses modèles de masquage sont différents, il faut faire une décomposition *sinusoïde + bruit.*

  Une composante tonale est un maximum local du spectre $S(k) > S(k+1)$ et $S(k) > S(k-1)$, tel que :

$$S(k) - S(k+j) \geq 7dB$$

$$j\epsilon[-2,+2] \text{ si } 2 < k < 63$$
$$j\epsilon[-3,-2,+2,+3] \text{ si } 63 \leq k < 127$$
$$j\epsilon[-6,\ldots,-2,+2,\ldots,+6] \text{ si } 127 \leq k \geq 250$$

  Il faut ajouter à son intensité celle de la composante d'avant et d'après. Les autres composantes tonales dans la même bande de fréquence ne sont plus considérées.

  Les composantes qui ne sont pas tonales sont composés par une somme des intensités des composantes du signal qui restent dans chacune des 24 bandes critiques comprises entre 0 et $15500Hz$. (Le système auditif se comporte comme un banc de filtres passe-bande, avec des différentes fréquences centrales. Ces "filtres auditifs" peuvent s'approximer par des filtres rectangulaires avec des largueurs de bandes critiques qui augmentent avec la fréquence. Dans ce modèle, la bande audible est divisée dans 24 bandes critiques non-réguliers).

- Troisième phase : Effacer les composantes masquées.

  Il faut éliminer les composantes situées en dessous du seuil absolu de masquage et les composantes tonales séparées moins de $0.5Barks$.

  Le Bark est une mesure de fréquence telle que $1Bark$ = largueur d'une bande critique:

$$\text{pour des fréquences } f < 500Hz,\ 1Bark \approx \tfrac{f}{100}$$
$$\text{pour des fréquences } f > 500Hz,\ 1Bark \approx 9 + 4log(\tfrac{f}{1000})$$

– Quatrième phase : Seuils de masquage individuels et globaux Dans cette phase, le système va tenir en compte l'effet de masquage fréquentiel du système auditif. Il faut discrétiser l'axe de fréquences selon la sensibilité auditive, qui est meilleure pour les basses fréquences et qui définie les fréquences en Barks. Les courbes de masquage sont donc plus ou moins linéaires (avec des pentes différentes selon la distance entre la composante masquée et la composante qui masque) et elles dépendent d'un indice de masquage différent pour les composantes tonales et pour les autres composantes. Le système utilise $f_1$ pour dénoter l'ensemble de fréquences présentes dans le signal de test. Le seuil global de masquage pour chaque fréquence $f_2$ tient en compte le seuil auditif absolu $S_a$ et les courbes de masquage $P_2$ des $N_t$ composantes tonales et $N_n$ composantes non tonales:

$$S_m(f_2) = 10log_{10}[10^{S_a(f_2)/10}$$

$$+ \sum_{j=1}^{N_t} 10^{P_2(f_2,f_1,P_1)/10}$$

$$+ \sum_{j=1}^{N_n} 10^{P_2(f_2,f_1,P_1)/10}$$

Le seuil de masquage est donc le minimum entre les maximums locaux de masquage et le seuil absolu dans chacune des 32 sous-bandes de largueur de bande constante dans le spectre.

Pour qu'il soit inaudible, le tatouage doit tomber en dessous du seuil de masquage.

# Annexe B

# Comparaison entre les différentes techniques de modulation

## B.1 Comparaison entre les différentes techniques de modulation binaires et quaternaires

Il y a été choisi comme modulation dans le système de tatouage une modulation de phase quaternaire QPSK d'une porteuse donnée. Dans ce chapitre, nous allons essayer de justifier le choix de ce système entre les différents systèmes de modulation possibles.

Il existe une grande quantité de schémas de modulation/détection possibles pour dessiner un système de communication numérique qui transmet des données dans un canal de transmission passe-bande. Le choix d'un schéma donné va être fait en fonction à certains paramètres qui sont :

1. Le débit : il s'agit de maximiser le débit de transmission maximum.
2. Minimisation de la probabilité d'erreur de symbole.
3. Minimisation de la puissance transmise.
4. Minimisation de la largueur de bande utilisée pour la transmission
5. Maximisation de la résistance aux signaux d'interférence.
6. Minimisation de la complexité du circuit.

Quelques objectifs sont contradictoires, par exemple 1 et 2 sont contradictoires à 3 et 4. Il faut donc arriver à un compromis.

Deux systèmes avec le même nombre de symboles peuvent être comparés seulement s'ils utilisent la même quantité d'énergie pour transmettre chaque bit d'information. Le coût de transmission est représenté par l'énergie totale nécessaire pour transmettre la totalité du message, et pas par la quantité nécessaire pour bien transmettre un seul symbole. Pour comparer les différents systèmes de modulation, le paramètre utilisé est la *probabilité d'erreur par symbole* exprimée en fonction du rapport signal à bruit en terme d'énergie $E_b/N_0$.

La table suivante présente les formules pour la probabilité d'erreur de symbole $P_e$ pour des modulations PSK cohérent, QPSK et MSK, quand la transmission s'effectue dans un canal $AWGN$:

| Modulation | Probabilité d'erreur, $P_e$ |
|---|---|
| PSK cohérent | $\frac{1}{2}erfc(\sqrt{E_b/N_0})$ |
| PSK cohérent différentielle | $erfc(\sqrt{E_b/N_0}) - \frac{1}{2}erfc^2(\sqrt{E_b/N_0})$ |
| FSK cohérent | $\frac{1}{2}erfc(\sqrt{E_b/2N_0})$ |
| DPSK | $\frac{1}{2}exp(-E_b/N_0)$ |
| FSK non cohérent | $\frac{1}{2}exp(-E_b/2N_0)$ |
| QPSK | $erfc(\sqrt{E_b/N_0}) - \frac{1}{4}erfc^2(\sqrt{E_b/N_0})$ |
| MSK | $erfc(\sqrt{E_b/N_0}) - \frac{1}{4}erfc^2(\sqrt{E_b/N_0})$ |

à partir de ces expressions, ainsi que de ses représentations graphiques, on arrive à certaines conclusions :

1. Les probabilités d'erreur $P_e$ diminuent quand la valeur de la relation $E_b/N_0$ augmente.

2. Pour toutes les valeurs de cette relation $E_b/N_0$, la modulation de phase PSK cohérent produit un taux d'erreur plus petit que tous les autres systèmes. Dans les systèmes de transmission binaire perturbés par un bruit blanc gaussien, la modulation optimale est donc la modulation PSK cohérent, dans le sens qu'elle minimise la probabilité d'erreur pour une relation $E_b/N_0$ donnée.

3. Les systèmes cohérents PSK et DPSK requièrent une relation $E_b/N_0$ $3dB$ plus petite que pour la modulation de fréquence FSK (cohérent ou non-cohérent), pour obtenir la même probabilité d'erreur.

4. Dans les valeurs hautes de $E_b/N_0$, DPSK et FSK non-cohérents fonctionnent si bien que le système PSK cohérent et le système conventionnel FSK cohérent, avec une différence de $1dB$, pour le même nombre de bits par seconde et la même énergie par bit.

5. Dans la modulation QPSK, on utilise deux porteuses orthogonales $\sqrt{\frac{2}{T}}\cos(2\pi f_c t)$ et $\sqrt{\frac{2}{T}}\sin(2\pi f_c t)$. La fréquence porteuse $f_c$ est un multiple entier du taux de symbole $\frac{1}{T}$. Le résultat est la transmission de deux chaînes de bits indépendantes. Pour des grandes valeurs de $E_b/N_0$, QPSK et PSK cohérent ont le même taux d'erreur.

6. Dans la modulation MSK, les deux porteuses orthogonales, $\sqrt{\frac{2}{T}}\cos(2\pi f_c t)$ et $\sqrt{\frac{2}{T}}\sin(2\pi f_c t)$, sont modulées par les deux pulses $\cos(\pi t/2T_b)$ et $sin(\pi t/2T_b)$ respectivement, dans des intervalles de durée $2T_b$, où $T_b$ est la durée de bit. Le récepteur utilise un processus de décodification de phase cohérent au cours de 2 bits successifs pour récupérer la chaîne de bits originale. La modulation MSK possède les mêmes performances que QPSK au niveau de probabilité d'erreur.

## B.2 Comparaison des différentes techniques de modulation M-aires

Dans cette section, nous essayons de comparer des différents schémas de modulation M-aires. Le tableau suivant indique des valeurs typiques de puissance et largueur de bande nécessaires dans une modulation M-aire en fonction de la valeur de $M$. Ce tableau montre que la modulation **QPSK** ($M = 4$) offre le meilleur compromis entre les besoins de puissance et largueur de bande. C'est pour cette raison que la modulation QPSK est la plus utilisée en pratique. Pour $M > 8$, le besoin de puissance devient excessif, et c'est pour cela que ces types de modulation ne sont pas très employés. Aussi, la génération de signaux devient très complexe dans les schémas de modulation cohérents M-aire PSK, sur tout pour $M > 8$, et ce type de modulation n'est pas non plus utilisé.

La modulation PSK M-aire et la modulation en quadrature QPSK M-aire ont à peu près les mêmes performances au niveau puissance et largueur de bande. Néanmoins, pour $M > 4$, les deux systèmes ont des différentes constellations de signaux. Pour la modulation PSK M-aire, la constellation de signaux est circulaire, tandis que pour la modulation QAM M-aire est rectangulaire. Une comparaison entre ces deux constellations de signaux nous montre que la distance entre les points du dictionnaire pour la modulation PSK M-aire est plus petite que celle de la modulation QAM M-aire, pour une puissance fixe. Dans un canal AWGN (bruit blanc centré additif Gaussien), les performances de la modulation M-QAM sont meilleures que celles de la modulation M-PSK pour $M > 4$. Néanmoins, cette supériorité se trouve seulement quand le canal n'a pas de non-linearités.

| Valeur de $M$ | $\frac{(LargueurdeBande)_{M-aire}}{(LargueurdeBande)_{Binaire}}$ | $\frac{(PuissanceMoyenne)_{M-aire}}{(PuissanceMoyenne)_{Binaire}}$ |
|---|---|---|
| 4 | 0.5 | $0.34dB$ |
| 8 | 0.333 | $3.91dB$ |
| 16 | 0.25 | $8.52dB$ |
| 32 | 0.2 | $13.52dB$ |

# Table des figures

# Bibliographie

[1] R. Dixon, *Spread-Spectrum Systems*, John Wiley and Sons, 1976.

[2] S. Haykin, *Communication Systems*, John Wiley and Sons, 1994.

[3] L. Boney, A. Tewfik, K. N. Hamdy, *Digital Watermarks for Audio Signals*, IEEE Int. Conf. on Multimedia Computing and Systems June 17-23, Hiroshima, Hiroshima, june 1996.

[4] L. Boney, A. Tewfik, K. N. Hamdy, *Digital Watermarks for Audio Signals*, Eurospeech, Rhodes, september 1997.

[5] P. G. Flikkema, *Spread-Spectrum Techniques for Wireless Communication*, IEEE Signal Processing Magazine, may 1997.

[6] J. Lacy, S. Quackenbush, A. Reibman, D. Shur, J. Snyder, *On Combining Watermarking with Perceptual Coding*, Proc. Int. Conf. Acoust. Speech and Signal Processing, Seattle, may 1998.

[7] T. Furon, P. Duhamel, *An Asymmetric Public Detection Watermarking Technique*, Workshop on Information Hiding, Dresden, september 1999.

[8] R. A. Garcia, *Digital Watermarking of Audio Signals using a Psychoacoustic Auditory Model and Spread Spectrum Theory*, 107th AES Convention, New York, september 1999.

[9] T. Furon, N. Moreau, P. Duhamel, *Audio Public Key Watermarking Technique*, Proc. Int. Conf. Acoust., Speech and Signal Processing, Istambul, june 2000.

[10] E. Allamanche, J. Herre, *Secure Delivery of Compressed Audio by Compatible Bitstream Scrambling*, 108th AES Convention, Paris, february 2000.

[11] C. Neubauer, J. Herre, *Audio Watermarking of MPEG-2 AAC Bitstreams*, 108th AES Convention, Paris, february 2000.

[12] L. de C.T. Gomes, M. Mboup, M. Bonnet, N. Moreau, *Cyclostationarity-based Audio Watermarking with Private and Public Hidden Data*, 109th AES Convention, Los Angeles, september 2000.

[13] Site de MPEG, *http://www.mpeg.org.*

[14] Site du SDMI, *http://www.sdmi.org.*

[15] Site de l'AES, *http://www.aes.org.*

[16] Site de l'OPIMA, *http://www.cselt.it/ufv/leonardo/opima.*

[17] Serveur de l'Union Européenne *http://europa.eu.int.*

[18] Site de l'Institut d'Etudes Politiques de Paris *http://www.droitconstit.org.*