الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique
Université Akli Mohand Oulhadj - Bouira -
Tasdawit Akli Muḥend Ulḥağ - Tubirett -

وزارة التعليم العالي والبحث العلمي
جامعة أكلي محند أولحاج
-البويرة-

جامعة البويرة

Faculté des Sciences et des Sciences Appliquées

كلية العلوم والعلوم التطبيقية

Référence : ...../MM/2021

المرجع:...../م م/ 2021

Mémoire de Master

Présenté au

Département : Génie Électrique

Domaine : Sciences et Technologies

Filière : Télécommunications

Spécialité : Systèmes desTélécommunications

Réalisé par :

NOUAR Oualid

Et

SAADI Said

Thème

Tatouage et stéganographie des signaux audio

Soutenu le: **...../..../2021**

Devant la commission composée de :

Mr :	CHELBI Salim	Prof.	Univ. Bouira	Président
Dr :	SAIDI Mohammed	M.A.A	Univ. Bouira	Rapporteur
Mr :	BOUZIDA	M.C.B	Univ. Bouira	Examinateur
	……..	M.A.A	Univ. Bouira	Examinateur

Année Universitaire: 2020-2021

Dédicaces

Je dédie Ce modeste travail :

A mes parents. Aucun hommage ne pourrait être à la hauteur de l'amour dont ils necessent de me combler. Que dieu leur procure bonne santé et longue vie.

A mes frères «**Ibrahim** et **Mohammed** ».

A tout mes amis « **Zin El AbidineBouchnak**,**FaresBouamra**,**Fares Adad**, **djamelchaibi**,**Elouanas**, **Kadiro**, **Charef**,**Mourad,Nassim Ben Akil**et **MokhtarDendani**».

A mon binôme**Said**. Et toute la famille**NOUAR**.

Et à tous ceux qui ont contribué de prés ou de loin pour que ce projet soit possible, je vous dis Merci.

walid

Remerciements

Tout d'abord, nous tenons à remercier " Dieu " le tout puissant, qui nous a donné la force, la patience et la volonté d’accomplir ce modeste travail. Sans oublie nos parents pour leur contribution, leur soutien, leur patience et leurs encouragements, Nous tenons à remercier **Dr.SAIDI Mohammed**, notre encadreur et enseignent au département génie électrique pour son encadrement, son suivi, sa disponibilité, ses conseils précieux et son encouragement. Nos remerciements vont aussi Aux membres de jury d'avoir accepté de juger et d'évaluer ce travail. Nous tenons à remercier en cette occasion tout le corps professoral et administratif de département génie électrique de l’Université Akli Mohand Oulhadj de Bouira pour la richesse et la qualité de leurs enseignements et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée. Nous adressons nos plus sincères remerciements à tous nos proches amis, qui nous ont toujours encouragée au cours de la réalisation de ce mémoire.

Résumé

L'émergence des fichiers numériques offre des possibilités innovantes dans le domaine de la technologie des données, les fichiers hôtes, dans notre cas, les signaux audio, peuvent masquer des informations numériques. Ces techniques d'enterrement s'appellées le tatouage et la stéganographie. Notre travail à mettre en place une procédure de tatouage et de stéganographie qui permet l'intégration de la protection du droit d'auteur pour des données audio numériques en modifiant directement les échantillons audio. Ce processus utilise directement le masquage de perception, de temps et de fréquence pour garantir que le tatouage et la stéganographie numériques sont inaudibles et robustes.

Mots clés : Tatouage audio, stéganographie audio, psycho-acoustique, Masquage perceptuel.

Table des Matières

Chapitre II : psycho-acoustiques et masquage perceptuel

ChapitreIII: Résultats et discussions

Liste des Figures

Liste des Tableaux

Listes des Acronymes

- **Acronymes**

AMR	Adaptive Multi-Rate
CD	Compact Disc
DCT	DiscreteCosineTransform
DWT	DiscreteWaveletTransform
DRT	Discrete Ripplet Transform
DSSS	Direct Sequence Spread Spectrum
DRS	Dynamic Regular groups Steganalysis
FFT	Fast Fourier Transform
FM	Frequency modulation
FHSS	Frequency Hopped Spread Spectrum
JTC	Joint TechnicalCommittee
IVDS	Interactive Vidéo and Data Services
ISF	Immittance Spectral Frequencies
IFFT	IrrevesibleFast Fourier Transform
IEC	International Electrotechnical Commission
ISO	International Organisation for Standardisation
LSB	Least Significant Bit
MPEG	Moving Picture Expert Group
MP3	MPEG-1/2 Audio Layer 3
MELP	Mix-Excitation LinearPrediction
PESQ	Perceptual Evaluation of Speech Quality
QIM	Quantization Index Modulation
SSC	Spread Spectrum Coding
SC	SubCommittee
SPL	Sound Pression Level
SD	Spectral Distortion
SVQ	Split VectorQuantizer

VCD	Vedio Compact Disc
VQ	VectorQuantization
WG	Working Group
WB	Wide-Band

Introduction Générale

Dans le monde numérique d'aujourd'hui, il existe une grande richesse d'informations accessibles qui diverses formes (texte, images, audio et vidéo).Il est facile d'assurer la sécurité des "documents analogiques" et de protéger l'auteur contre le vol ou la copie de son œuvre.

La question est de savoir comment protéger les droits d'auteur ou étiqueter les informations numériques et préserver leur sécurité sans détruire ni modifier le contenu d'information.Le tatouage et la stéganographie été présentées comme des moyens efficaces de communication secrète.

Le tatouage est une application qui embarque une petite quantité de données mais est nécessaire pour la protection du droit d'auteur [34].Une approche de la sécurité des données consiste à utiliser le cryptage [34], cependant, une fois les documents déchiffrés, le tatouage ou "signature" est supprimée et il n'y a aucune preuve de propriété telle qu'une étiquette, un tampon ou un tatouage.

Le tatouage doit: être inaudible [34-35], être statistiquement invisible pour empêcher la détection et/ou la suppression non autorisée par des «pirates», avoir des caractéristiques de compression similaires à celles du signal d'origine pour survivre aux opérations de compression/décompression, être robuste aux attaques délibérées de « pirates », être robuste aux opérations de manipulation et de traitement de signal standard sur les données de l'hôte, par exemple le filtrage, le rééchantillonnage, la compression, le bruit, le recadrage, les conversions A/D-D/A,….etc.

La stéganographie audio peut transférer secrètement des messages importants en les incorporant dans des fichiers audio de couverture à l'aide de techniques de dissimulation d'informations [32]. La dissimulation des données dans l'audio est particulièrement difficile car le système auditif humain fonctionne sur une plage dynamique plus large par rapport au système visuel humain.

ChapitreI:
Généralités sur le tatouage et la stéganographie audio

I.1Introduction :

De nos jours, l'amélioration significative d'internet permet d'accéder facilement à différentes données multimédias (audio, vidéo, image). Ainsi, divers types de nouveaux défis liés à la protection du droit d'auteur et à la trempe du contenu sont introduits chaque jour. Le tatouage et la stéganographie numérique ont été utilisés efficacement pour relever ces nouveaux défis. Il s'agit d'un processus d'intégration d'informations secrètes dans des contenus numériques pour plus d'authenticité. Les principales applications du tatouage numérique comprennent l'authentification des données, la prise d'empreintes digitales, la protection des droits d'auteur, la protection de la propriété et la surveillance de la diffusion [3].

la stéganographie est l'art de cacher des informations secrètes dans un support donné (couverture), de sorte que le support résultant (stégo) soit quasiment identique au support donné. La notion de dissimulation de données ou de stéganographie a été introduite pour la première fois avec l'exemple du message secret des prisonniers par Simmons en 1983 [23].

Dans ce premier chapitre, nous avons présenté le tatouage et la stéganographie numérique appliquer sur un signal audio. La figure suivante représente les différentes techniques de la sécurité de l'information en général.

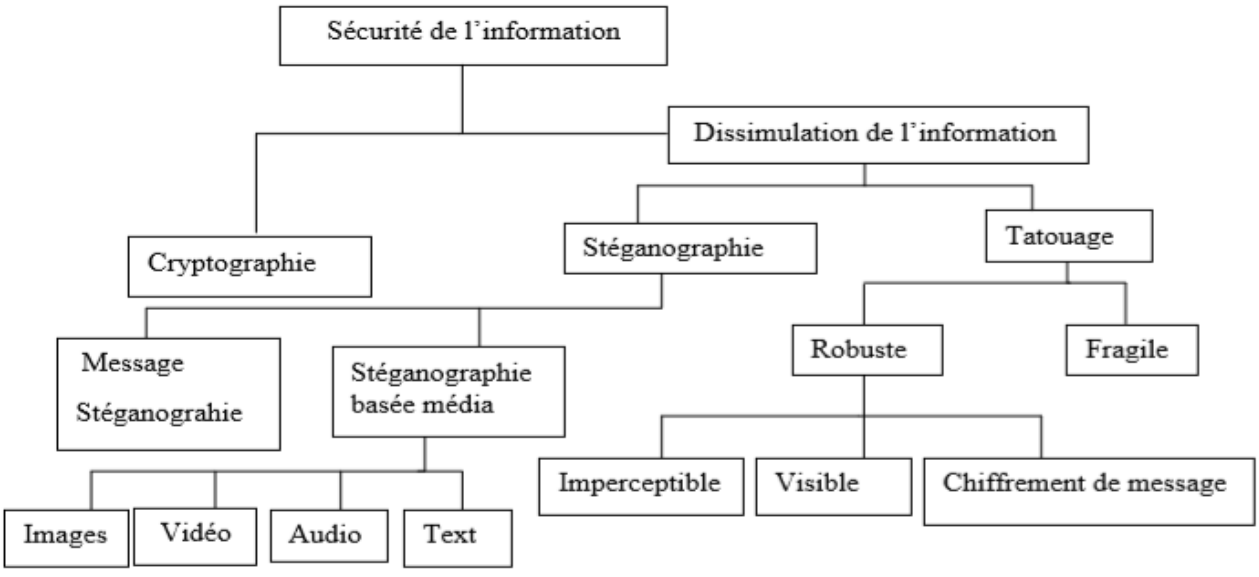


Figure I.1: Techniques de la sécurité de l'information [8].

I.2 Tatouage et stéganographie

Le tatouage fait généralement référence à l'insertion d'informations dans un contexte de sécurité, notamment pour prévenir le piratage : les informations généralement associées à une signature ou à une marque sont associées au contenu du document lui-même. Elle régit la propriété, l'authenticité et les droits d'utilisation des documents [17,18]. Contrairement à la stéganographie, elle suppose que la communication est connue d'un tiers (hacker). Le tatouage a pour but d'ajuster la signature en fonction du comportement du pirate sur le document : si l'on veut lutter contre les documents piratés, la signature doit être indélébile ; au contraire, si l'on veut garantir l'intégrité du document, il doit être immédiatement après que le pirate modifie le document disparaissent [14]

I.3 La dissimulation de l'information

Le masquage de données comprend le masquage d'informations dans un document composé de données numériques. En dehors du domaine numérique, ces pratiques sont très anciennes et ne sont plus utilisées aujourd'hui. Dans la dissimulation d'informations, le problème classique de la communication avec des données cachées a été proposé pour la première fois par Simmons. Il existe deux principes pour l'insertion d'informations. La stéganographie recouvre la communication, cette action assure la confidentialité des données transmises. Un autre principe est le tatouage, qui permet d'insérer une information dans le document et de l'authentifier et son intégrité est garantie [20].

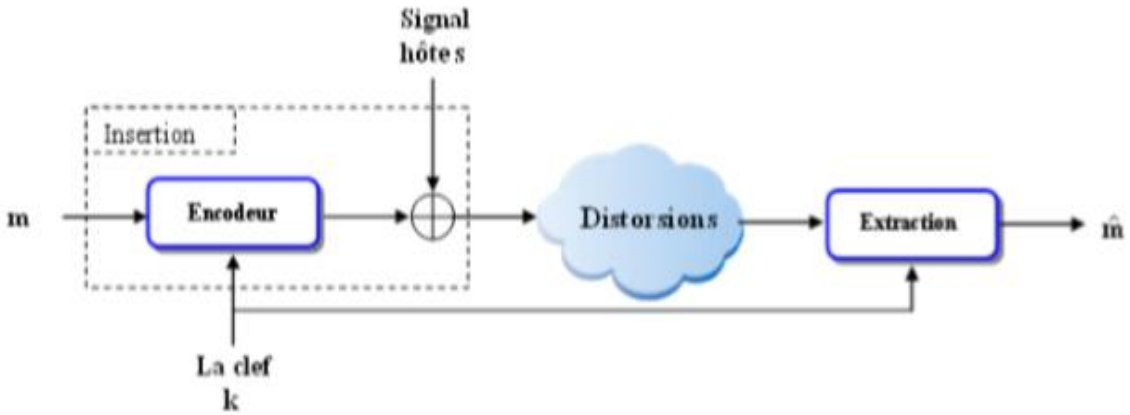


Figure I.2: Schéma général d'un système de dissimulation d'information [19].

I.3.1 Le schéma de dissimulation

L'ensemble du processus de masquage des informations repose sur deux opérations :

- ✓ **Insertion :** comprend l'insertion d'informations dans un support (document original dans un tatouage ou support de couverture en stéganographie).

- ✓ **Extraction :** Qui obtient cette information.

Lorsqu'il s'agit de vérifier l'existence d'informations dans un support tatoué, la détection de mots est également utilisée au lieu de les extraire [20].

I.3.2 Les caractéristiques de dissimulation d'information

Les applications de dissimulation sont triées en fonction de trois critères [20] :

- ✓ **L'invisibilité:**Les données ne doivent pas être visibles sur les supports stéganographiques. Le but du tatouage n'est pas d'endommager le document protégé, c'est-à-dire que la marque doit être invisible.
- ✓ **La Capacité :** La capacité d'un système caché est le rapport de la quantité de données à cacher sur la taille du support (support couvert). Ou la capacité est le nombre de bits valides cachés dans le support de couverture.
- ✓ **Robustesse :** C'est la résistance à diverses dégradations ou attaques. Ces attaques sont appelées "attaques aveugles" car le pirate informatique ne sait pas réellement ce qu'il fait.

De manière générale, la capacité est faible, plus la robustesse et l'invisibilité sont fortes. Si une marque de grande taille est insérée, le document tatoué risque d'être bien dégradé et le tatouage perd son invisibilité [20].

I.3.3 Les différentes techniques de dissimulation d'information

Les algorithmes de masquage d'informations peuvent être distingués les uns des autres des quatre manières suivantes [20] :

- ✓ Méthode de sélectionner les points ou blocs dans le fichier de couverture (objet de couverture) qui portent les informations à masquer.
- ✓ Le choix de l'espace de travail pour les opérations d'insertion (domaine spatial, domaine de transformation, tel que DCT, DWT, Fourier Merlin, etc.).
- ✓ Le principe de mise en forme des informations à masquer avant insertion (redondance, code correcteur d'erreur, etc.).
- ✓ Méthode d'insérer des informations dans un média de couverture (objet de couverture).

I.4 Tatouage ou « Watermarking »

Le tatouage d'un signal comprend l'insertion et le masquage d'informations binaires dans le signal, de façon imperceptible, cela comprend également l'insertion d'une signature ou d'informations inaudibles dans le signal audio. Cette information représente une séquence pseudo-aléatoire. La détection du tatouage est effectuée en utilisant la corrélation entre le tatouage d'origine et l'estimation du tatouage à partir du signal d'observation [14].

I.4.1 Tatouage numérique

Le tatouage numérique est une sorte de marqueur secrètement intégré dans un signal tolérant au bruit tel que des données audio, vidéo ou d'image. Il est généralement utilisé pour identifier la propriété du droit d'auteur de ce signal [1]. Le tatouage numérique n'est apparu que très récemment, au milieu des années 90. Il consiste à insérer une signature imperceptible et ineffaçable dans le document [4].

Les marques sont généralement insérées dans le « domaine spatial ou fréquentiel ». Cependant, la marque insérée doit respecter deux restrictions fondamentales : imperceptible et indélébile. L'imperceptibilité signifie que la distorsion doit être suffisamment faible pour que l'utilisateur ne puisse pas distinguer le document tatoué du document original. Indélébile signifie que la marque insérée ne peut pas être effacée après diverses attaques bien intentionnées ou malveillantes. Cette dernière nécessite des contraintes de robustesse, car tant que le document est disponible, la signature doit y rester, c'est-à-dire que la marque ne peut être effacée ou rendue inutilisable sans altérer la qualité du document, et il n'y a qu'une déformation évidente. Si l'attaque est exécutée, elle devrait apparaître [5].

Dans ce manuscrit nous allons nous intéresser au tatouage des documents de type audio.

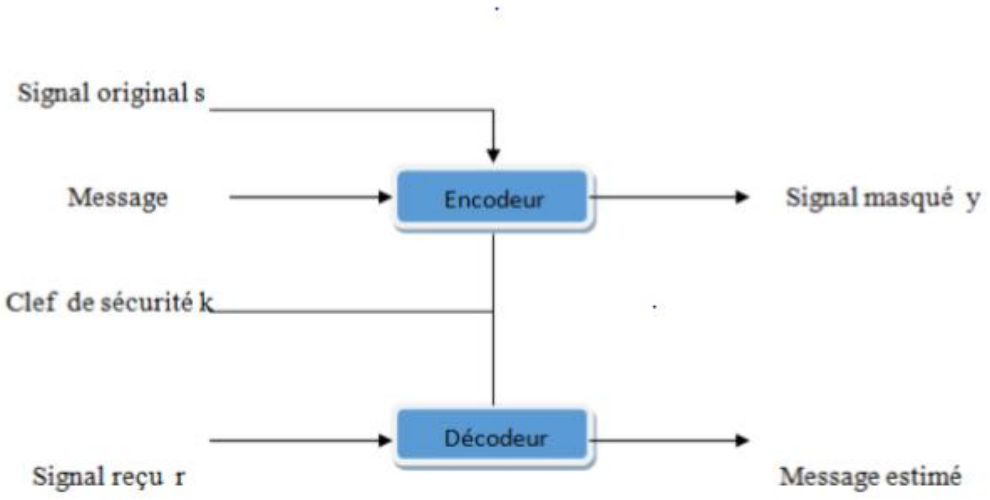


Figure I.3 : Schéma général d'un système de tatouage numérique [19].

I.4.2 Tatouage audio

Les signaux audios sous forme numérique sont très faciles à reproduire. Par conséquent, une technologie de protection efficace (telle que le tatouage) devient essentielle pour vérifier l'identité de l'auteur du document. La présence de la marque ne doit pas provoquer de distorsion auditive. Dans ce cas, les défauts du système auditif humain sont utilisés pour ajouter des informations

inaudibles au son. Et pour les traitements habituels du signal (filtrage, compression, etc.) et les attaques malveillantes, il doit également être robuste. Cependant, ce signe ne peut être détecté que par le personnel autorisé [6].

I.4.2.1 Applications du tatouage audio

Les applications attenantes aux techniques de tatouage pour la transmission de données peuvent permettre :

- **L'annotation de documents sonores :** pouvant servir d'aide à l'indexation dans les bases de données : L'information cachée a` 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 [9].
- **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 [10].
- **L'identification du document :** Cette application, souvent associée au tatouage de protection, peut être envisagée dans un contexte de transmission de données d`es 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 [11].
- **Le contrôle d'applications cible :** L'information peut être destinée à une application adjacente à l´écoute du signal audio, nécessitant la mise en œuvre d'un décodeur spécifique [12].
- **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 (Interactive Vidéo and Data Services) a d'ailleurs été proposé : l'information cachée contient la référence du produit mis en vente par le télé-achat [13].

I.4.2.2 Objectifs et contraintes du tatouage audio

Les systèmes de tatouage, s'ils sont utilisés dans des applications d'amélioration de contenu, doivent répondre à un ensemble spécifique d'objectifs et de contraintes. Avoir :

- **L'inaudibilité du tatouage :** les informations tatouées ne doivent pas réduire de manière perceptuelle la qualité du signal audio qui y est inséré, son insertion doit être transparente [14].
- **Le débit et la fiabilité de transmission :** La mission du système est de transmettre des informations par des signaux audio. Il doit fournir la plus grande fiabilité de transmission possible à un taux de transmission aussi élevé qu'une chaîne de communication traditionnelle [14].

- **Robustesse :** la transmission des informations cachées dans le signal audio suppose que le système mis en œuvre soit robuste aux interférences introduites par tous les canaux de diffusion par lesquels le signal audio peut être transmis : si la dégradation affecte le signal audio tatoué, la transmission doit être maintenue fiabilité. Contrairement aux tatouages protecteurs, ces perturbations ne sont pas affectées par le concept de pirates. Étant donné que les informations de tatouage sont destinées à augmenter le contenu de l'utilisateur ou de l'application cible, les pirates n'aideront pas en bloquant sa détection. Ainsi, les interférences à considérer correspondent à toutes les dégradations légales qui peuvent être apportées au signal audio lors de la production en studio ou lors de la transmission dans le réseau de diffusion [15,16].
- **Capacité d'insertion :** la quantité maximale d'informations pouvant être transmises lorsque la probabilité d'erreur est quasi nulle [14].
- **Détection de tatouage aveugle :** La détection d'informations cachées doit être effectuée directement à partir du signal audio avec le tatouage, sans connaître le signal audio d'origine [14].
- **Coût :** Certaines applications étant conçues pour fonctionner en temps réel, le coût en termes de temps de calcul peut être le critère déterminant pour le choix d'une technologie de tatouage [14]

I.5 Stéganographie

Le terme stéganographie est dérivé des mots grecs « stégos » signifiant « couverture » et « grafia » signifiant « écriture » la définissant comme « écriture couverte » [21]. La stéganographie est la pratique consistant à cacher des informations « à la vue de tous ». Cette technique repose sur l'encodage et le cache d'un message dans une couche de transport de manière à rendre l'existence du message inconnue d'un observateur [22].

1.5.1 Cryptage et stéganographie d'information

En cryptographie, la structure d'un message est brouillée afin de le rendre insignifiant et inintelligible, sauf si la clé de décryptage est disponible. Elle ne tente pas de déguiser ou de cacher le message codé. Fondamentalement, la cryptographie offre la possibilité de transmettre des informations entre des personnes d'une manière qui empêche un tiers de les lire. La cryptographie peut également fournir une authentification pour vérifier l'identité de quelqu'un ou de quelque chose [30].

En stéganographie, on ne modifie pas la structure du message secret, mais on le cache à l'intérieur d'une image de couverture de manière à ce qu'il ne soit pas visible. Un message dans un texte chiffré, par exemple, peut éveiller les soupçons du destinataire, alors qu'un message "invisible" créé à l'aide de méthodes stéganographiques ne le fera pas. En d'autres termes, la

stéganographie empêche un destinataire involontaire de soupçonner l'existence des données. En outre, la sécurité du système stéganographique classique repose sur le secret du système d'encodage des données. Une fois le système d'encodage connu, le système de stéganographie est mis en échec [31, 32].

I.5.2 Caractéristiques de Schéma Stéganographique

Trois critères sont utilisés pour classer les algorithmes stéganographiques : capacité, invisibilité et robustesse [33].

- **Capacité :** Correspond à la taille des données pouvant être incorporées dans l'objet de couverture, par rapport à sa taille.
- **Invisibilité ou transparence :** dépend directement de la distorsion introduite par le processus de masquage dans le processus d'insertion de données, la distorsion n'est que la quantité de modification ou de changement dans l'objet couvert.
- **Robustesse :** Cela signifie la résistance de notre stégo-objet, c'est-à-dire qu'il reste normal même s'il subit une conversion (filtrage, etc.).

Ces trois critères ne peuvent pas être maximisés simultanément. Chacun d'entre eux aura une influence sur l'autre. Par exemple, la capacité va en contradiction avec la transparence

I.5.3 Les types de la stéganographie

Il existe trois types de la stéganographie peut être divisée en :

I.5.3.1 Stéganographie pure

Lorsque le système n'a pas besoin de modifier les informations secrètes, il est appelé une « stéganographie pure », telle que stégo-Key. Le processus peut être représenté comme suit : mappage $E : C \times M \rightarrow C$, où$C$ représente toutes les couvertures possibles, et M est tous les messages possibles. Alors que la formule $D : C \rightarrow M$ est utilisé pour extraire tout message secret existant hors de la couverture, bien sûr $|C| \geq |M|$ est une condition nécessaire. Les algorithmes ici ne sont pas publics, seuls les expéditeurs ou les destinataires peuvent accéder au processus d'extraction et d'intégration [24].

Ce type offre le moins de sécurité, car la communication qui l'utilise conduit les expéditeurs et les destinataires à ne dépendre que de l'hypothèse selon laquelle les autres parties ne se soucient pas des messages secrets qu'ils envoient, comme les messages Internet [25].

I.5.3.2 Stéganographie à clé privée (secrète)

Ce type, il est nécessaire de changer la clé secrète via la communication. Le message est intégré dans un message de couverture à l'aide d'une clé secrète (stégo-key). Les personnes qui

connaissent la clé secrète peuvent obtenir le message original qui se trouve à l'intérieur du message de couverture et le lire. Alors qu'en stéganographie pure, la clé secrète est plus susceptible d'être interceptée, en stéganographie privée, il existe un canal secret et des échanges de clés stego à travers celui-ci. En cas d'interception de messages stéganographiques privés, seules les personnes connaissant les clés peuvent voir le message et l'extraire [25].

I.5.3.3 Stéganographie à clé publique

Ce type s'appuie sur les moyens des clés cryogéniques publiques, qui sont des systèmes à la fois publics et secrets pour sécuriser les canaux de communication entre les personnes qui souhaitent communiquer en privé. Côté envoi, le message est codé par clé publique, mais seule la clé secrète qui correspond mathématiquement à ce code public peut décoder le message.

Les avantages de la stéganographie à clé publique se trouvent clairement dans la mise en œuvre plus efficace de la cryptographie à clé publique, elle comporte différentes étapes de sécurité, qui empêche d'autres parties de se joindre à la communication sans se méfier de la stéganographie utilisée, et elles doivent également savoir comment déchiffrer l'algorithme utilisé [25].

I.5.4La stéganographie appliquée sur les signaux audios

La meilleure technique d'adaptation peut être choisie après avoir compris par quels moyens le signal audio est transféré :

1.5.4.1 Modification d'amplitude

Une méthode populaire de modification de l'amplitude est appelée insertion de LSB, cette méthode peut être mise en œuvre dans la stéganographie ou le tatouage [24]. Il est évident que dans cette méthode, les données sont codées dans les LSB des données audios. Par exemple, prenons un fichier échantillonné sur 16 bits, le masquage des données peut utiliser les quatre bits les moins significatifs. Néanmoins, les données cachées peuvent subir des déformations et être détectées alors qu'elles sont cachées. Les données cachées peuvent être alertées par une ressemblance ou un bruit. Mais les changements dans le LSB peuvent rendre le bruit audible [26]. Risque de distorsion des données lors de la copie, de la compression ou de la conversion A/D, D/A [27].

L'insertion des LSB est une technique simple, mais elle est rarement utilisée, ce qui rend les autres techniques plus sympathiques.

1.5.4.2 Codage de phase

Une autre méthode de dissimulation des données audio est le codage de phase, qui fait référence à la phase du signal audio par rapport à une phase de référence, qui représente les données. Toutes les phases sont ajustées pour correspondre à la phase relative requise.

Le codage de phase est efficace pour le codage des signaux afin d'ajuster le rapport de bruit. Lorsque les phases entre les segments de fréquence changent, une dispersion évidente apparaît. De toute façon, des modifications inacceptables peuvent être détectées, dans le cas de petites modifications de la phase. Dans ce cas, un codage inaudible peut être effectué.

En général, les auditeurs ne peuvent pas remarquer de changements sur le son lorsque le codage de phase est effectué avec des déphasages lisses, ce qui fait du codage de phase l'une des méthodes les plus efficaces en termes de rapport de bruit [27].

Le codage de phase a un débit de données de soutien très faible. En fait, c'est l'un des principaux inconvénients de cette méthode [28].

I.5.4.3 Codage du spectre étalé

Une autre méthode pour cacher des informations dans les signaux audios est le "Spread Spectrum Coding" SSC. La conservation de l'énergie et de la largeur de bande est une nécessité dans le transfert des données audio, donc le canal de communication tend à rendre la région de fréquence des signaux audio aussi étroite que possible.

Sur [29], une méthode discutée nommée codage "Direct Sequence Spread Spectrum" (DSSS). Dans cette méthode, une puce est utilisée pour multiplier un signal audio afin de l'étaler, une séquence plus longue est modulée à un certain taux. La fréquence d'échantillonnage peut être choisie comme la fréquence de la puce, car les signaux de l'hôte sont déjà des signaux discrets. Un des problèmes de l'utilisation des DSSS est de déterminer le début et la fin des quanta de verrouillage de phase de la puce, ce qui doit être pris en considération par la nature des signaux discrets. Cependant, on dispose d'un plus grand débit de la puce et donc d'un plus grand débit de données associées. En fait, sans cela, de nombreux algorithmes de verrouillage du signal pourraient être utilisés, mais ils sont coûteux. Malheureusement, le DSSS entraîne un bruit aléatoire supplémentaire dans le signal audio, et le codage électronique ne permet pas d'obtenir des résultats satisfaisants [29].

La méthode de codage "Frequency Hopped Spread Spectrum" (FHSS) utilise également le spectre étalé. Dans cette technique, la fréquence du signal porteur passe rapidement d'une certaine fréquence à une autre lorsqu'il est alerté d'une certaine manière [24]. Le nouveau signal audio est divisé en petites parties, chaque partie étant portée par une fréquence spéciale qui lui est associée [24].

L'utilisation du codage à étalement de spectre est principalement utile en raison de sa résistance aux variations. Il est difficile d'ajuster les données intégrées sans faire de destruction observable pour les données de couverture. Ainsi, les données sont étalées à travers les informations de couverture.

1.5.4.4 Masquage des données d'Echo

Dans cette méthode de codage, un signal audio est considéré comme un signal hôte dans lequel les données sont intégrées, en fournissant un écho. Le masquage des données s'effectue en ajustant trois paramètres de l'écho : le taux de décroissance, l'amplitude initiale et le décalage. Lorsque le décalage entre le signal nouveau et l'écho diminue, les deux signaux se combinent. A un certain point, l'homme ne peut plus observer de différence entre ces deux signaux. L'écho est ici introduit comme une résonance supplémentaire [29].

Les données peuvent être cachées dans le fichier audio en utilisant des délais dissemblables qui séparent le signal nouveau et le signal écho. Le signal nouveau peut être divisé en plusieurs petites parties, afin de permettre l'intégration de plus d'un bit, chaque partie de ce signal peut être répercutée sur un certain bit. Tous les segments codés indépendants existent dans les dernières données de couverture [24].

Cette technique "Echo Data Hiding", fonctionne particulièrement bien pour les fichiers sonores qui ne contiennent pas de dégradation ajoutée, comme le cas où il y a un temps de silence dans le fichier, ou quand il y a des pertes dans l'encodage ou du bruit dans le signal [26].

I.6 Conclusion

Le tatouage est un procédé assez récent, inspiré de techniques anciennes. La stéganographie peut cacher le maximum d’information utile dans un signal audio, ainsi, l’existence d’informations secrètes dans les signaux stégos est c’est difficile à détecter, presque impossible à détecter.Dans ce chapitre, nous avons introduit de sujet lié au tatouage et stéganographie appliquer sur les signaux audios.

Chapitre II:
psycho-acoustiques et masquage perceptuel

II.1Introduction :

La contrainte imposée au tatouage d'un signal audio est celle del'inaudibilité ; le tatouage est supposé être imperceptible. Cela conduit à une question cruciale qui est la suivante : comment peut- on embarquer une information dans un signal hôte sans qu'elle soit audible ?

Il est alors nécessaire de faire appel à la psychoacoustique pour établir la condition d'inaudibilité.

Dans ce chapitre, nous décrivons l'utilisation d'un modèle perceptuel adéquatà notre travail.

Nous exploitons le principe des masquages fréquentiel et temporel quisont le noyau de notre procédure permettant de rendre inaudible la signature.

II.2 Psychoacoustique

II.2.1 Présentation

La psychoacoustique est l'étude des rapports entre les paramètres de la stimulation acoustique d'une part, et la qualité de la sensation auditive, d'autre part. Elle comprend l'étude de la représentation dans le système nerveux 'codage nerveux' des différentes dimensions des stimuli acoustiques, ainsi que l'étude du type et des règles de fonctionnement mis en jeu par les systèmes auditifs pour structurer la perception, y compris les mécanismes d'ordre supérieur appelés « processus cognitif » [36].

En fait, la transmission du message acoustique nécessite des signes chargés de signification, ils sont au nombre de trois : l'intensité, la fréquence et le temps. L'intensité et ses variations étant d'importance minime, c'est l'association fréquence-temps qui représente le vecteur le plus efficace [36].

La psychoacoustique s'intéresse généralement à 4 types de perception : la perception de l'intensité sonore, la perception de la hauteur tonale, la détermination des seuils, l'audition binaurale et notamment la fonction de localisation auditive. La perception auditive a son niveau d'organisation de la perception de la parole et notamment la distinction du signal utile du reste de l'environnement sonore [36].

II.2.2 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 que si son niveau d'intensité est supérieur 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 suivante [37] :

$$S_E(f) = 3.64f^{-0.8} - 6.5\exp[-0.6(f-3.3)^2] + 10^{-3}f^4 \ \ (dB\ SPL) \quad \text{(II.1)}$$

f désigne la fréquence en Khz

II.2.3 Bandes critiques

L'oreille peut être vue comme un analyseur fréquentiel du signal sonore, cela peut être modélisé par un banc de filtres passe-bande, appelés filtres auditifs dont les bandes passantes, appelées bandes critiques, se recouvrent en se chevauchant sur la gamme des fréquences audibles (de 20 Hz à 20 kHz), elles sont définies autour d'une fréquence centrale pour laquelle la largeur de bande est augmentée jusqu'à ce qu'il y ait une différence apparente dans la tonalité.

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, qui démontrent que les largeurs des bandes sont non-uniformes et augmentent avec la fréquence, car le système auditif possède une bonne résolution spectrale en basses fréquences mais médiocre en hautes fréquences. Les expériences réalisées 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 [37].

Ce banc de filtres caractérise efficacement le mode de perception de l'oreille, son pouvoir de séparabilité et sa résolution fréquentielle.

Le tableau suivant montre la position des bandes critiques dans le domaine fréquentiel :

N°dela bande	Fréquencecentral (Hz)	Bornes de la Bande
1	50	-100
2	150	100-200
3	250	200-300
4	350	300-400
5	450	400-510
6	570	510-630
7	700	630-770
8	840	770-920
9	1000	920-1080
10	1175	1080-1270
11	1370	1270-1480
12	1600	1480-1720
13	1850	1720-2000
14	2150	2000-2320
15	2500	2320-2700
16	2900	2700-3150
17	3400	3150-3700
18	4000	3700-4400
19	4800	4400-5300
20	5800	5300-6400
21	7000	6400-7700
22	8500	7700-9500
23	10500	9500-12000
24	13500	12000-15500
25	19500	15500-

Tableau II.1: Fréquences des bandes critiques selon [37]

Toutes les expériences réalisées donnent, à peu de différences, lesmêmes fréquences et positionsdes bandes critiques .

II.2.3.1Echelle de Bark

Une nouvelle échelle des fréquences, celle des Bark, a été définie pour faciliter les traitements dans les bandes critiques. L'unité perceptive Bark, assure le lien entrela fréquence d'un son exprimé en Hertz et la résolution de l'oreille, un Bark correspond à la largeur d'une bande critique. L'expression analytique la plus utilisée pour rendre compte de la correspondance entre unefréquence et un taux de bande critique Bark est celle de Zwicker [38] :

$$f_{bark} = 13\ arctang\left[\frac{0.76\ f_{hertz}}{1000}\right] + 3.5\ arctang\left[\left(\frac{f_{hertz}}{7500}\right)^2\right] \text{(II.2)}$$

Le tableau suivant représente la correspondance entre Hertz et Bark [38] :

Bark	Hertz	Bark	Hertz
1	100	**13**	2000
2	200	**14**	2320
3	300	**15**	2700
4	400	**16**	3150
5	510	**17**	3700
6	630	**18**	4400
7	770	**19**	5300
8	920	**20**	6400
9	1080	**21**	7700
10	1270	**22**	9500
11	1480	**23**	12000
12	1720	**24**	15500

Tableau II.2 :CorrespondanceHertz-Bark .

L'échelle Bark est approchée par unbanc de 24 filtres triangulaires espacés linéairement jusqu'à1Khz, puis espacés logarithmiquement jusqu'auxfréquences maximum.

La psycho-acoustiques'attache à modéliser le système auditif humain et la perception du son. Ses résultats, essentiellement basés sur des expérimentations, mettent en évidence les propriétés du système auditifet plus particulièrement les phénomènes de masquage.

Plusieurs modèles psychoacoustiquesont utilisés dans le tatouage numérique audio, notamment le modèle de Gomes [39], celui de Garcia [40] et lemodèle MPEG n°1 [41] que nous

utilisons dans notretravail.

II.2.4 La compression audio et la psychoacoustique

La compression audio se différencie de la compression vidéo par :

- les volumes de données en jeu (pour une même durée).
- les principes sur lesquels elles reposent.

Si on peut résumer la compression vidéo par le principe :

« Ne jamais transmettre une donnée déjà transmise »

La compression audio pourrait se résumer par :

« Ne jamais transmettre un son inaudible »

La compression audio repose complètement sur des études psycho-acoustiques et la connaissance du système auditif humain.

II.3 Masquage perceptuel :

II.3.1 Présentation :

Un son en présence d'un autre peut devenir partiellement ou complètement inaudible : c'est ce que l'on appelle l'effet de masque.

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.

L'effet de masquage dépend des caractéristiques spectrales et temporelles des signaux masqué et masquant. La procédure que nous utilisons exploite directement les caractéristiques du masquage temporel et fréquentiel pour embarquer un tatouage inaudible et robuste.

II.3.2 Masquage fréquentiel :

Le masquage fréquentiel se réfère au masquage entre des composantes fréquentielles dans un signal audio. Quand deux signaux apparaissent simultanément et se rapprochent fréquentiellement, le plus fort d'entre eux peut rendre le plus faible inaudible [14].

Le système auditif a tendance à ne prendre en compte que le son le plus fort et négliger le reste. Ainsi, si une tonalité faible se situe dans la bande critique d'une tonalité plus forte, la tonalité faible ne sera pas perceptible.

Le seuil de masquage d'un son masquant dépend de la fréquence, du niveau de pression du son (SPL) et des caractéristiques des composantes tonales (sinusoïdes) ou non tonales (bruits) des deux sons masquant et masqué [42]. La nature des sons, sinusoïdaux ou bruités, joue 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 masquant et masqués [14].

La figure suivante montre une fonction d'étalement dont le signal masquant est un bruit de bande centré sur 1 kHz de largeur 160 Hz présenté aux niveaux LG = 100, 80, 60, 40 et 20 dB , le signal masqué est une sinusoïde[14] .

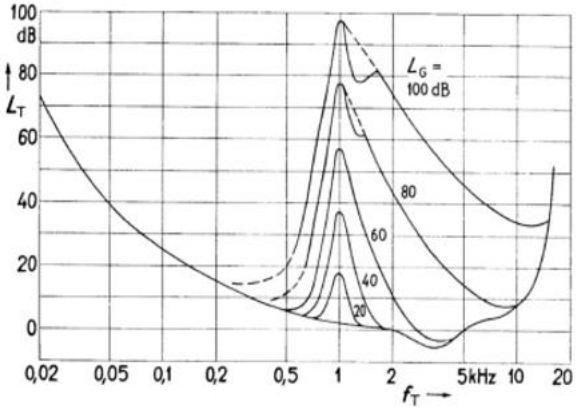


Figure II.1: Fonctions d'étalement pour des bruits à bande étroite[14].

Il faut préciser qu'il est plus aisé pour un bruit à bande étroite de masquer une sinusoïde (indice de masquage ~ 4dB), que pour un signal tonal de masquer un bruit à bande étroite (indice de masquage~ 24dB). Par ailleurs, les signaux à haute fréquence sont plus facilement masqués, cela est dû à la résolution fréquentielle du système auditif vue précédemment.

II.3.2.1 Seuil de masquage

Le seuil de masquage fréquentiel est une courbe dans le domaine fréquentiel au-dessous de laquelle tout son inséré est masqué.

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, 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.

Les modèles de masquage fréquentiel sont le plus souvent obtenus des codecs audiocourants de haute qualité. Dans ce travail, nous utilisons le modèle défini dans la norme 'ISOMPEG Audio Psychoacoustique Model 1, Layer 1.

II.3.2.2 Norme ISO/MPEG

II.3.2.2.1 Présentation

L'ISO (International Organisation forStandarisation) est le plus grand producteur et éditeur mondial de Normes internationales. L'ISO est un réseau d'instituts nationaux de normalisation de 161 pays, selon le principe d'un membre par pays, dont le Secrétariat central, situé à Genève, Suisse, assure la coordination d'ensemble [43].

L'ISO se charge des normes dans des domaines aussi variés que le traitement de l'information et les communications, les textiles, la production d'énergie, la distribution des biens, les services financiers, etc. Au sein de l'ISO, le groupe WG11 de la sous-commission29(ISO/IEC/JTC1/SC29/WGll), connu sous l'acronyme MPEG (Moving Picture Expert Group) a pour mission le développement des normes multimédias basées sur l'audio et la vidéo numérique [44].

Les normes MPEG de l'ISO sont des Normes internationales qui traitent de la compression, de la décompression, du traitement et de la représentation codée de l'image animée, du son et de leur combinaison. Il existe plusieurs normes publiées dans ce domaine, auxquelles on se réfère couramment par les sigles MPEG 1, MPEG 2, MPEG 4, MPEG 7 et MPEG 21 [43].

La norme MPEG est une norme propriétaire et protégée, d'où la difficulté d'accéder aux spécifications et détails techniques de cette dernière.

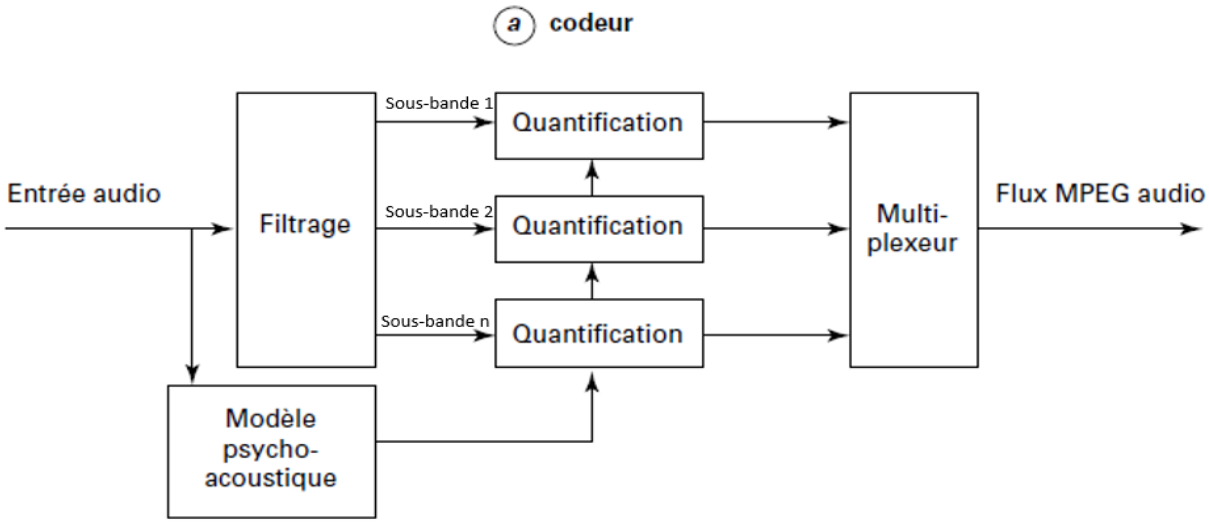
II.3.2.2.2 Compression MPEG audio :

La base algorithmique de la compression audio MPEG est le système acoustique humain, qui n'a pas les mêmes caractéristiques qu'un instrument d'enregistrement. Comme vu précédemment, l'oreille humaine est un système non linéaire à seuillage adaptatif. En premier lieu, ce seuillage (non-sensibilité à certains sons en deçà d'une puissance donnée) est variable en fonction de la fréquence, le maximum de notre sensibilité se situant en général entre 2 et 5 kHz. Ce modèle est compliqué par un phénomène de masquage. En effet, notre oreille percevra certains niveaux sonores

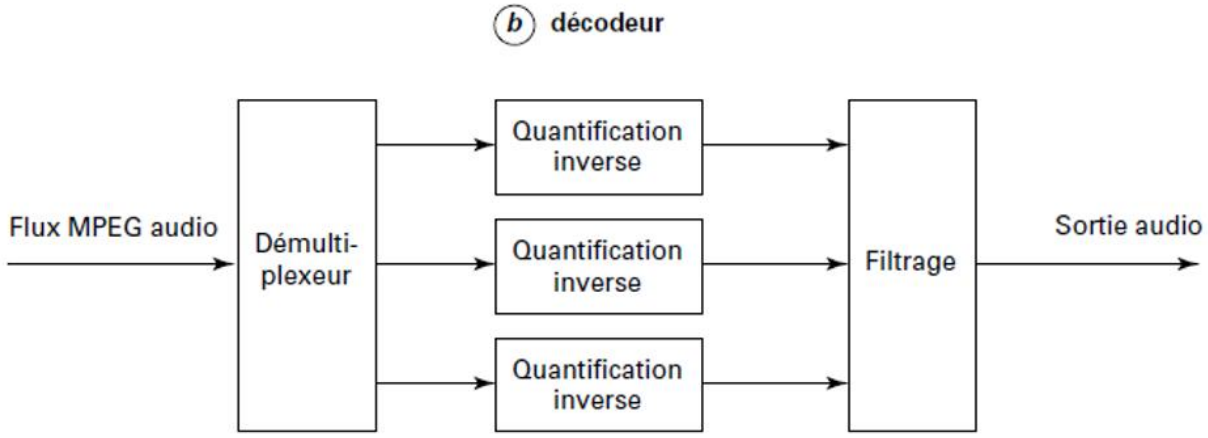
assez bas dans un silence total, alors qu'un signal sonore comportant des fréquences similaires masquera l'audibilité des mêmes sons. Le mode de compression MPEG met donc à profit ces caractéristiques pour dédier la bande passante numérique aux sons audibles par une oreille humaine.

L'objectif de prise en compte des caractéristiques auditives de l'oreille est réalisé parl'utilisation d'un mode de codage par sous-bandes (figures 2 et 3). Pour chaque sous-bande, le signal numérique d'entrée est traité par un filtre spécifique qui permet d'obtenir la composante sur cette bande de fréquence. Ensuite, chaque signal représentant la sous-bande est quantifié avec un pas dépendant du niveau de seuillage de la fréquence traitée. Le processus qui détermine le pas de quantification pour chaque sous-bande fait appel au modèle psychoacoustique. Le choix de ce modèle détermine la qualité du codeur ainsi que sa complexité, les autres fonctions se retrouvant à l'identique dans chaque codeur audio. Cette opération permet de supprimer dans le signal les informations les moins perçues par l'oreille humaine. On transmet donc dans le flux MPEG les valeurs quantifiées ainsi que le pas de quantification utilisé dans chaque bande de fréquence. Le décodeur, après démultiplexage des données, quantification inverse et filtrage inverse, pourra reconstituer le signal décodé [45].

Le flux audio MPEG est organisé en trames contenant un nombre fixe d'échantillons d'entrée (384 ou 1152). Aucune correspondance n'existe entre la durée des images vidéo et les trames audio. Au début de chaque trame, on trouve un en-tête avec un mot de signalisation et les informations de haut niveau nécessaires au décodage de la trame : fréquence d'échantillonnage du signal d'entrée, débit de sortie compressée, mode de codage utilisé. On trouve ensuite les valeurs du signal d'entrée après filtrage et quantification [45].



FigureII.2: CodageMPEG en sous-bandes audio [45].



FigureII.3:décodageMPEG en sous-bandes audio [45].

II.3.2.2.3 MPEG-1

L'application visée par la norme MPEG-1 est l'enregistrement. Elle a trouvé sa mise en œuvre dans plusieurs produits : le CD (Compact Disc), le VCD (Vidéo Compact Disc). Les premiers systèmes de télévision numérique aux États-Unis ont même fait appel à cette norme avant l'avènement de l'MPEG-2.

Pour l'audio, les fréquences d'échantillonnage autorisées sont 32Khz, 44.1Khz et 48 kHz. Les débits varient entre 32 et 384 kbit/s. Trois niveaux (layers) de codage audio sont utilisés. Ces niveaux se distinguent par les outils de compression, les fréquences d'échantillonnage du signal d'entrée et les débits [45].

- **Niveau 1 :** Débit de 384Kbit/s et un taux de compression de 3.23 .
  Le filtrage d'entrée est de type DCT avec utilisation d'un modèle psychoacoustique uniquement en fréquence.
- **Niveau 2 :** Débit de 160 à 256 Kbit/s et un taux de compression de 7.75 à 4.84 .
  Le filtrage d'entrée est aussi réalisé dans le domaine temporel, ce qui permet un certain masquage temporel.
- **Niveau 3 :** Débit de 112 à 128 Kbit/s et un taux de compression de 11.1 à 9.7 .
  Le filtrage d'entrée est modifié pour obtenir des largeurs de bandes de fréquences inégales et donc mieux adaptées au système auditif humain. Pour le cas d'un codage de signal stéréo, la cohérence entre les deux sources est utilisée. MPEG-1 audio niveau 3 est plus connu sous l'appellation MP3, qui est souvent déformée à tort en MPEG-3.

Ces valeurs ne signifient pas grand-chose sur la qualité du résultat obtenu, puisque la qualité ne dépend pas seulement du format de codage du fichier, mais également de la qualité de l'algorithme psychoacoustique utilisé par le codeur. Typiquement, les codeurs layer I utilisent un algorithme simple, d'où un résultat nécessitant un débit supérieur pour un codage transparent [46].

II.3.2.2.4 Modèle psychoacoustique MPEG-1, Niveau 1

Parmi les modèles psychoacoustiques développés dans le cadre de MPEG, nous nous sommes plus particulièrement intéressés au modèle psychoacoustique n°1. Ce modèle se base 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 ainsi le seuil de masquage global. Ce seuil est ensuite modifié pour prendre en compte le seuil d'audition absolu.

II.3.3 Masquage temporel

Le terme de masquage temporel fait référence au masquage réalisé après l'apparition d'un son masquant de forte intensité, mais également, ce qui peut paraître plus surprenant, au masquage avant la perception de ce son (dû à l'inertie de l'oreille). Le terme masquage 'temporel' indique que le son masquant et le son masqué sont décalés dans le temps, par opposition au masquage fréquentiel qui ne concerne que des sons simultanément présents. Les phases des sons masqué et masquant expriment le décalage temporel.

II.3.3.1 Pré-masquage et post masquage

Le masquage temporel peut donc 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. Le pré-masquage se produit entre 5 et 20 msec avant l'apparition du son masquant tandis que le post-masquage se produit entre 50 et 200 msec après la disparition du son masquant [42].

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, de la fréquence des deux sons, de l'intensité et de la durée du son masquant. La figure suivante illustre ce seuil [44].

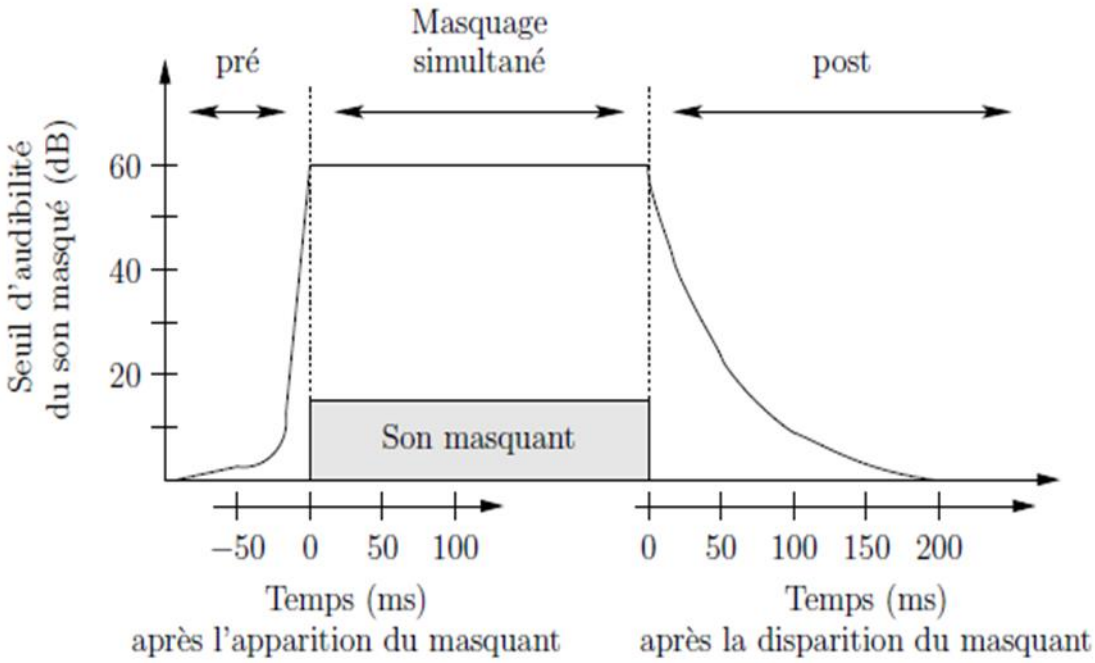


Figure II.4: Seuild'inaudibilitédusonmasquéenfonctionde l'ordreetdutemps d'apparition du son masqué et du son masquant.

Le masquage temporel et en particulier le post-masquage, n'est pas encore bien cerné, ses modèles associés sont difficiles à mettre au point [47].

L'effet du masquage temporel peut être approxime en utilisant l'enveloppe du signal hôte.

L'enveloppe est modélisée comme une décomposition en exponentiel. En particulier, l'estimation de l'enveloppe $t(n)$ d'un signal $s(n)$ accroit et décroît en e^{-at}en suivant l'évolution du signal [48].

Nous utiliserons dans notre travail la transformée d'Hilbert qui est adaptée à la modélisation de l'enveloppe temporelle d'un signal.

II.3.3.2 Transformée d'Hilbert

La transformée d'Hilbert est un outil mathématique très utilisé en théorie du signal, pour décrire l'enveloppe complexe d'une grandeur réelle modulée par un signal. Elle permet d'effectuer une démodulation fréquentielle ou d'amplitude. Il en résulte une enveloppe temporelle.

La transformée d'Hilbert peut aussi être décrite comme une séparation de la structure fine et de l'enveloppe d'un signal, sachant que la structure fine représente la parole, alors que l'enveloppe représente la hauteur [49].

L'expression de la transformée d'Hilbert est déduite à partir de la transformée de Fourier inverse du spectre d'un signal analytique associé à un signal physique, elle se présente comme suit [49] :

$$H[X(t)] = 1/\pi * X(t) \quad \text{(II.3)}$$

Où $X(t)$ est un signal et $(*)$ est une convolution.

La transformé de Hilbert de $X(t)$ et la convolution entre $X(t)$ et le signal $1/\pi t$. C'est la sortie filtrée de X(t) passant dans un filtre linéaire invariant dans le temps et ayant pour réponse impulsionnelle $1/\pi t$.

Un signal analytique est alors décrit comme suit :

$$\alpha X\ (t) = X(t) + j\ H[X(t)] \quad \text{(II.4)}$$

Sa partie réelle représente le signal physique [49].

II.3.3.1 Properité

- La transformée d'Hilbert est une transformation qui se limite au domaine temporel, contrairement à la transformée de Fourier.
- Le filtre de Hilbert est un filtre qui est considéré comme un quadrateur parfait, linéaire et dont la réponse impulsionnelle est $1/\pi t$. Il n'est donc pas causal, donc non réalisable physiquement. Il n'a d'intérêt que si le signal est à bande étroite [50], ce qui dans notre cas ne nous pose pas de problème.
- L'enveloppe résultante est le plus souvent un signal complexe de basses fréquences.

II.4 Conclusion

Le deuxieme chapitre nous avons parlé sur le model psycho-acoustique système auditif humain, pour comprendre les différentes transformations du signal audio pendant la transmission de l'information, après nous avons présenté lemasquagetemporel etlemasquage fréquentiel, puis, nous avons parlé surla méthode de compression audio « MPEG » qui est couvrent un très grand nombre d'applications.

Chapitre III:

Résultats et discussion

III.1 Introduction :

L'objectif de ce travail pour le tatouage est de présenter un schéma de tatouage d'audio applicable dans un contexte de communications numériques. Le cadre de ce travail nécessite des contraintes précises :

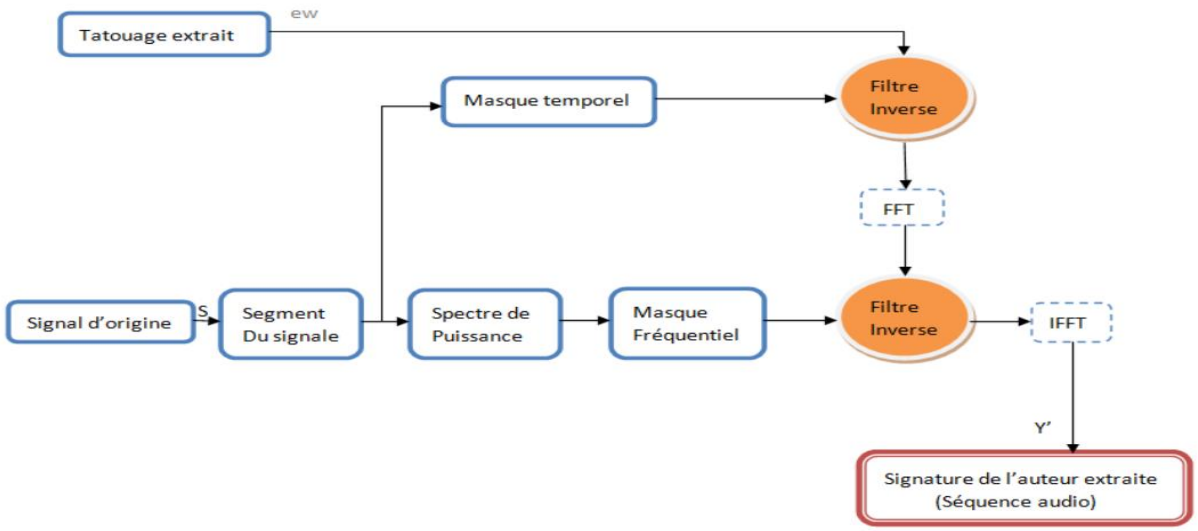
• Le débit de transmission du message inséré doit être le plus haut possible.

• Le système de tatouage doit également être robuste face à la compression de données, en particulier à la compression de type MPEG qui est souvent utilisée.

• Les attaques destinées à enlever la marque dans le signal de parole (piratage) ne seront pas prises en compte. Ceci car les signaux à transmettre ne sont pas des données sensibles. D'après une étude détaillée sur le tatouage, nous avons pu tirer que le domaine fréquentiel est un bon espace de point de vue robustesse et inaudibilité d'où l'idée d'utiliser la FFT et IFFT pour passer du domaine temporel au domaine fréquentiel et la transformation la plus utilisée dans le codage audio vue qu'elle permet une résolution fréquentielle plus fine tout en étudiant les effets de bordure. D'autre part, pour diminuer l'audibilité de la marque lors de l'insertion des bits de tatouage, nous avons exploité les propriétés du modèle psycho acoustique.

Pour la stéganographie nous présentons les différentes implémentations des techniques de stéganographie mises au point en utilisant les deux codeurs de parole AMR-WB G.722.2 et le MELP V 1.2, fonctionnant respectivement à 12.65 kbps et 2.4 kbps.

Lafigure suivante représente le shemagéniral du tatouage.



FigureIII.1: Schéma géniral du tatouage.

III.2 Etapes de l'implémentation

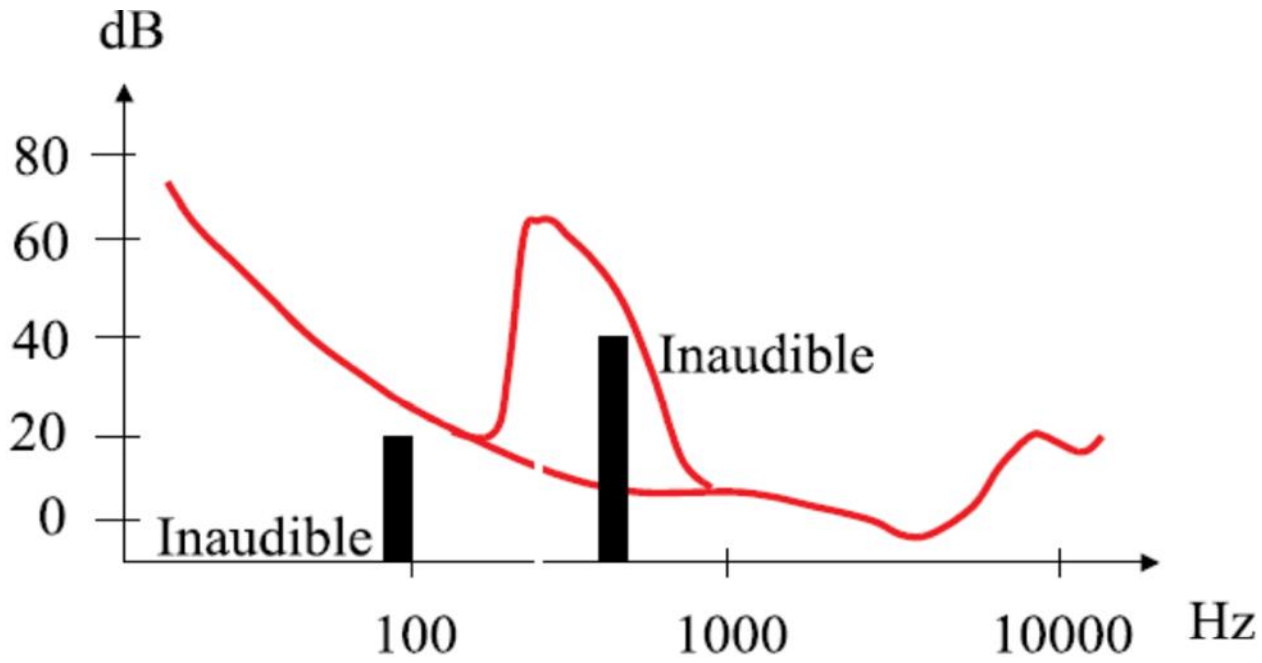
La mise en forme du signal à tatouer fait la diminution du volume des fichiers audio (et donc du débit nécessaire pour leur lecture), une première étape dans la chaîne de compression consiste donc à éliminer tous les signaux dont l'amplitude se situe en dessous du seuil de perception. Une séquence pseudo-aléatoire (la clé privée) est ajoutée au signal audio. Afin de garantir l'inaudibilité du tatouage, la séquence est mise en forme spectralement suivant un seuil de masquage obtenu à l'aide d'un modèle psycho acoustique. La détection est réalisée par une mesure de corrélation entre le tatouage original et un tatouage estimé à partir du signal observé. Cette détection est évidemment privée [52].

III.3 Détermination du masque fréquentielle

Ce phénomène étudié notamment par Fletcher et Zwitcker, met en évidence l'effet suivant : unson pur relève le seuil d'audibilité ou voisinage de sa fréquence. Il masque les sons d'intensité plus fable, situés à des fréquences proches de la sienne : voir la figure (III.2). On ne parlera plus alors de seuil d'audibilité, mais de seuil de masquage.

Les sons masqués constituent donc l'information non-pertinente, ils peuvent être éliminés du flux audio.

Cette propriété n'est pas typique des sons pues, mais elle s'applique également à un son quelconque, à condition de découpé son spectre en bande fréquentielle et d'analyser le masque auteur de chaque bande.



FigureIII.2: Masque fréquentielle [53].

Fletcher, puis zwitcker en met en évidence l'existence de bande fréquentielle critique appeler Bark ils en ont dénombré 24 sur l'ensemble du domaine audible. Tous les sons émis dans une mêmebande critique contribuent dans la même manière au masquage des bandes critiques voisines.

Par conséquence un bon modèle psycho acoustique devrait commencer par une décomposition du spectre en bande critique de $Bark$. Cependant, ceci compliquerait notablement l'implémentation, car ces bandes critiques n'ont pas des largeurs identiques, ni en valeur absolue, ni même en valeur relative, en moyenne et en haut fréquence les bandes critique sont fort mais pas en basse fréquence[53].

III.4 Identification des composantes tonale et no tonale

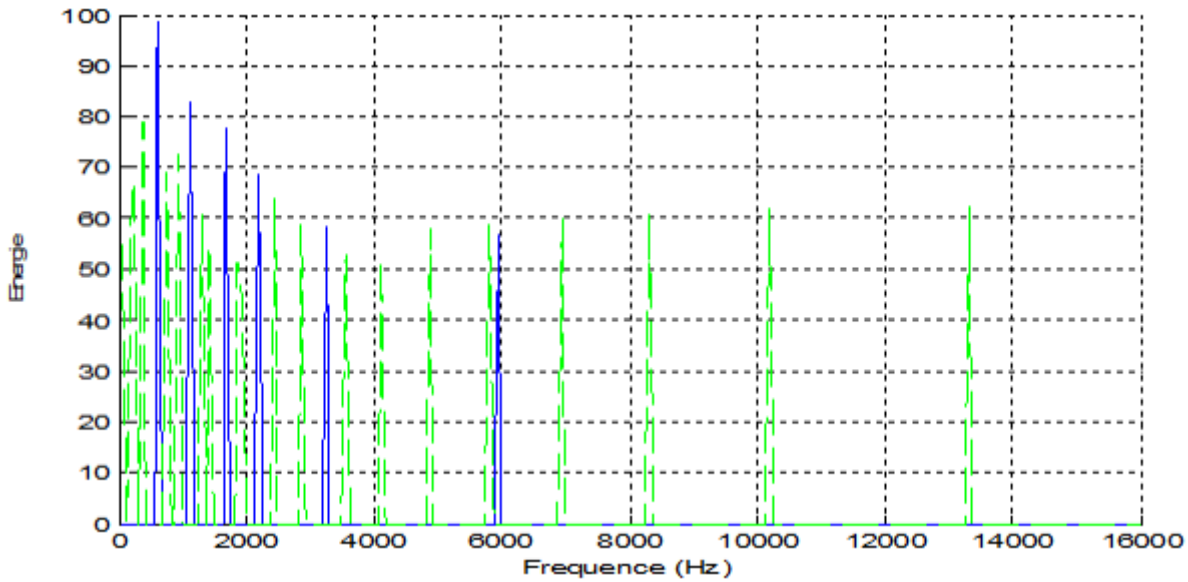
Dans le modèle MPEG, l'identification des composantes tonales et non tonales est un processus sélectif en fréquence. Un son pur, désigné dans le jargon de la psycho acoustique par le terme de tonale, généré une pression acoustique sinusoïdale dans le temps. Le niveau acoustique de ce son est représenté sur une échelle logarithmique. Il a l'allure d'une seule raie en fréquence. Un son pur estrarement rencontre dans la nature, une tonale au centre d'une bande critique va masquer tout bruit présent dans la même bande et d'intensité plus faible.

Les composantes tonales sont définies comme les maximums locaux du spectre de puissance c'est la composante est une tonale si elle est supérieure à ses voisines immédiates (maximum local) [54].

Nos ajoutons à son intensité celles du composant précédant et suivant d'autres composants tonalsdans la même bande de fréquence ne sont plus considérés. La composante non tonale est

constituée de la somme des intensités de la composante du signal restant chacune des 24 bande critiques entre 0 et $15500Hz$. Le système auditif comporte une bande de filtres passe-bande avec des fréquences centrales continuellement ouvertes ce filtre auditif peut être approché par des filtres rectangulaires avec une fréquence.

Dans ce modèle, la bande sonore est par conséquent divisée en 24 bandes critiques non régulières, les composantes tonales et non tonales du signal audiod'exemplesont représentées sur la figure (III.3)

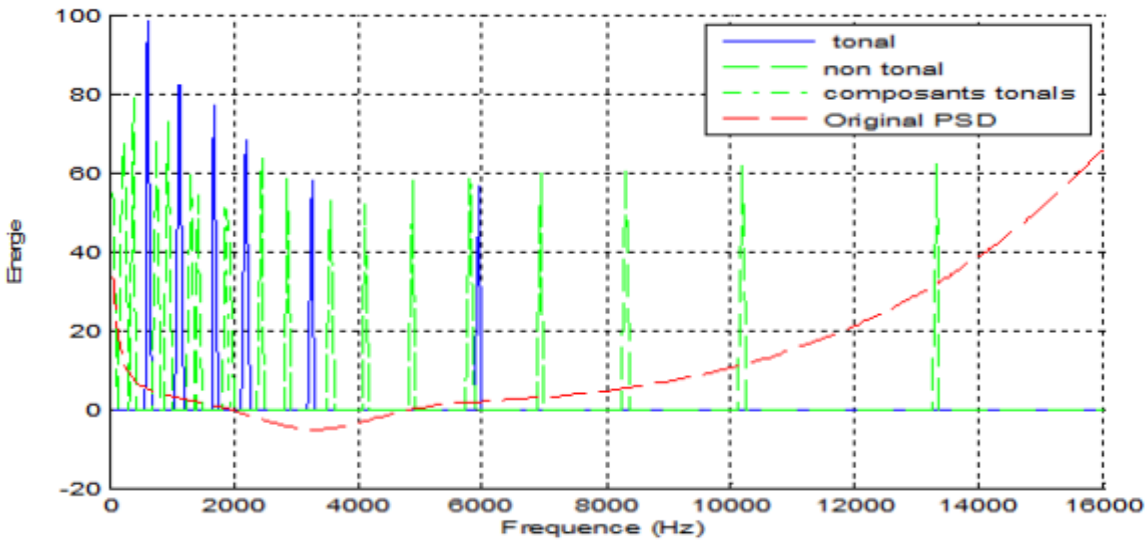


FigureIII.3 : Les composants tonals.

III.5Elimination des sons masque

Dans la télécommunication et pour transmis les donnes entre l'utilisateur ce dernier fait une zone des fréquences bien déterminer pour éliminer toutes les fréquences qui sont supérieur à une référence programmable appeler seuil de masquage à l'ide d'un threshold : le seuil (en dB sur le signal d'entrée) au-dessus duquel le signal est compressé.

Le composant doit être faible, le seuil d'audition absolue et le composant tonal séparé moins de 0.5 barks sont élimination (le bruit), et un complot des composants enlevés un long avec le seuil d'audition absolue est montré dans la figure (III.4).

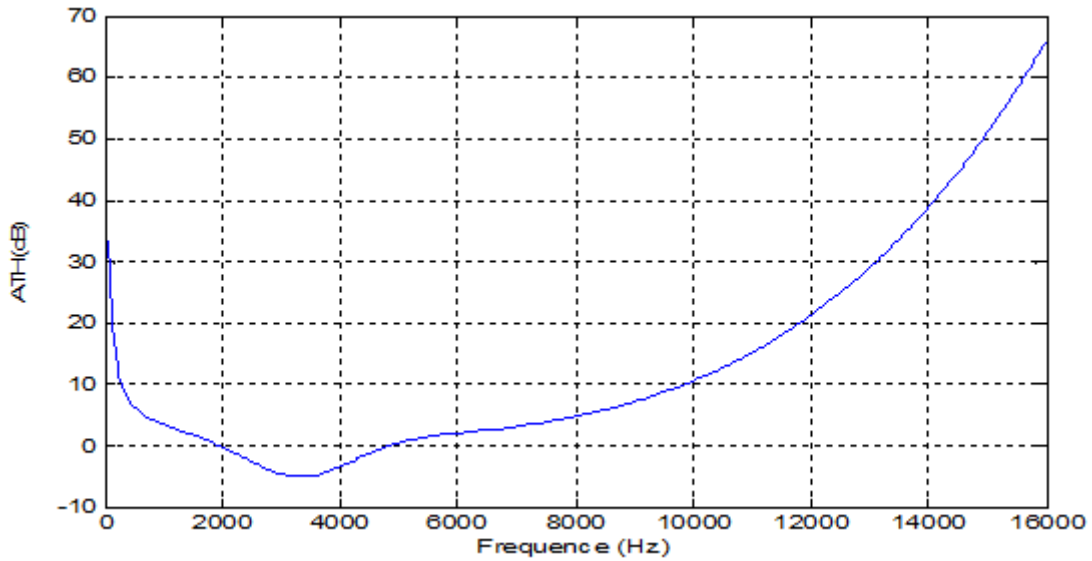


FigureIII.4 :élimination des composants masqués.

III.6 Détermination du seuil de masque fréquentiel

Ce seuil représente la limite supérieure du bruit pouvant être ajouté au signal original par les opérations de quantification, sans que ce bruit devienne audible, elle présente par la relation de :

$$ATH(dB) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5\exp\left[-0.6\left(\frac{f}{1000} - 3.3\right)^2\right] + 10^{-3}\left(\frac{f}{1000}\right)^4 \quad \text{(III.1)}$$



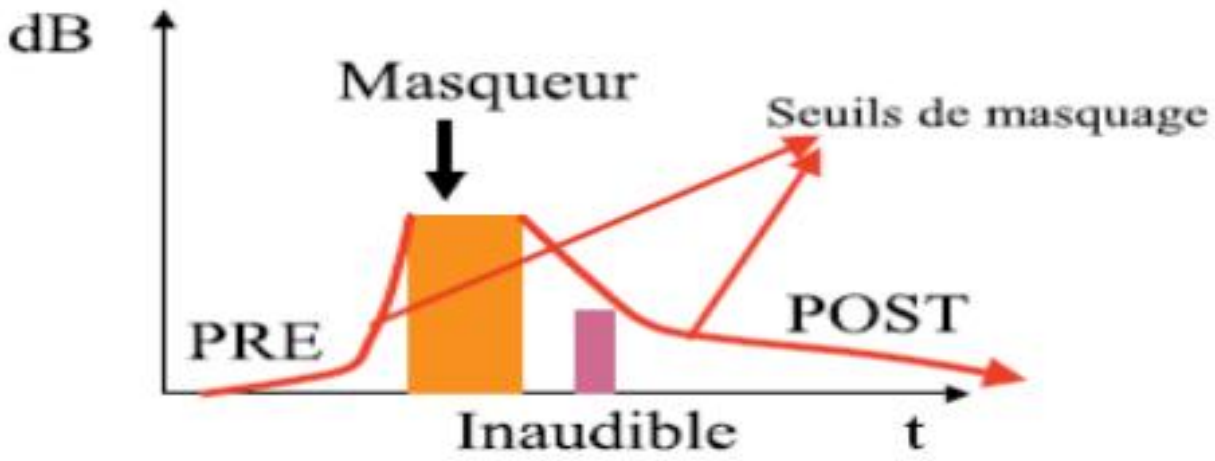
FigureIII.5 Seuil audition fréquentielle.

III.7 Détermination du masque temporel

On parle de masque temporel lorsqu'un son de faible puissance est masqué par l'apparition à un instant différent d'un autre son de forte puissance alors d'après les études qui nous allons vue

enpeut expliquer ce phénomène : un son de forte intensité masquer les sons d'intensité plus faible, immédiatement postérieure (poste masking) et antérieure (per masking), la figure (III.6) montre queles propriétés sont ici asymétriques : le pré-masquage concerne un intervalle de temps de quelque milliseconde seulement, alors que le post-masquage s'étend dans plusieurs dizaines de milliseconde.

Le phénomène de pré-masquage semble à première vue contredire le principe causalité, cependant, il ne faut pas oublier que notre oreille à un temps de réponse non nul, et qu'elle intègre, donc les stimulé sur un période correspondant à quelque $30msec$ [55].



FigureIII.6 : Masque temporel [55].

III.8 Génération de la signature de l'auteur

Le principe de cette technique est d'insérer directement dans un signal audio, une marque voireune signature, contenant des informations particulières comme le titre d'une chanson ou le nom d'un artiste, alors la signature de l'auteur c'est une séquence binaire qui est ajouté par l'utilisateurau signal audio comme un bruit aléatoire, elle prendre la même forme de signal audio transmis mais avec un principe c'est qu'à une faible puissance et faible amplitude para pore au signal original. On souhaite insérer la signature suivante :

$$Br = bruit\ (N/2) \qquad \text{(III.2)}$$

Cette signature est tout d'abord sur-échantillonnée avec un facteur d'équation global afin de créer un signal redondant, permettant de le rendre plus robuste avec cette multiplication :

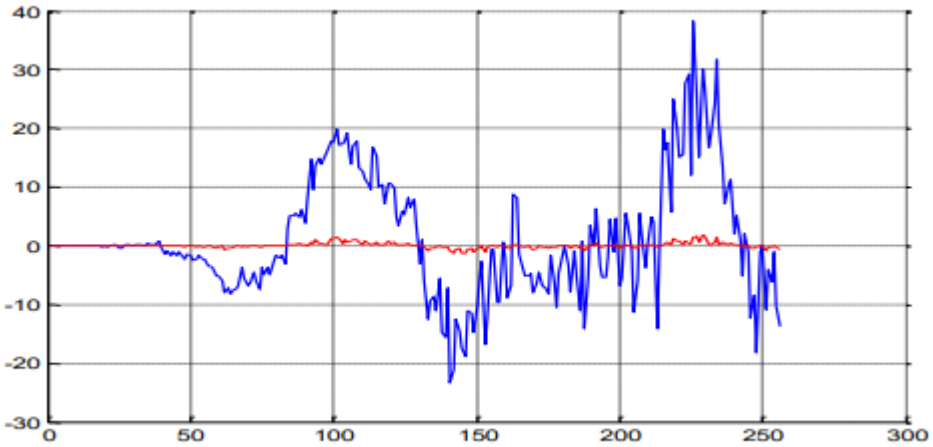
$$m = \ treshs * Br \qquad \text{(III.3)}$$

Après la transformer de Fourier inverse avec la valeur absolue de cette signature ont modulé ensuite le signal obtenu par un bruit binaire pseudo-aléatoire, qu'on amplifie ensuite par le multiplexage avec le filtre actif.

$$B = |ifft(m)| \qquad \text{(III.4)}$$

En revienne à l'insertion de la signature, et le tatouage est ajouté de la façon suivante :

$$x_2 = x_1\left(1{:}\frac{N}{2}\right) + B \qquad \text{(III.5)}$$



FigureIII.7 : Principe de signature par apport au signal audio.

III.9 Masquage de la signature

Cette signature est masquée à l'aide du seuil de masquage, pour extrairel'information non-pertinente, les échantillons dont l'intensité est située sous le seuil de masquage seront éliminés, de plus, les échantillons restants seront en codés avec moins $16bits$.

En codés avec moins de bits, cela revient à augmenter le bruit. Si l'on se rappelle l'analyse de rapport signal à bruit d'un encodeur. On tire de ce qui précède le principe de l'allocation des bits disponibles : le nombre de bit par échantillon doit être choisi de telle manière que le bruit reste inférieur de seuil de masquage (inaudible).

Un autre principe pouvant être utilisé pour réduire encoure le nombre de bit est le suivant : lors le passage du signal audio le niveau de bruit (inaudible, masqué), n'est pas élevé, et donc la réduction du nombre de bit ne sera pas très importante.

On peut alors amplifier leur de l'encodage les niveaux du signal les plus faibles, ce qui permettra d'accepter un bruit, les niveaux du signal élevé reste eux inchangés : on a donc opéré une compression de la dynamique du signal.

Lors du décodage, l'opération inverse sera effectuée (expansion de dynamique) : les niveaux les plus faible seront atténués du même facteur que lors de la compression, le même que le bruit qui passera sous le seuil de masquage [56].

III.10 Pondération avec le masque fréquentiel

L'analyse du phénomène de masquage lisse clairement sous-entendre qu'une analyse fréquentielle du signal est nécessaire, en particulier les premiers essais de réduire de bruit dans les années 1960ont montré que la compression large bande était de mauvaise qualité, cela suggère qu'adopter un bruit.

Par conséquent il est nécessaire, de travailler par bandes de fréquence. Mais avec quelles largeurs de bande.

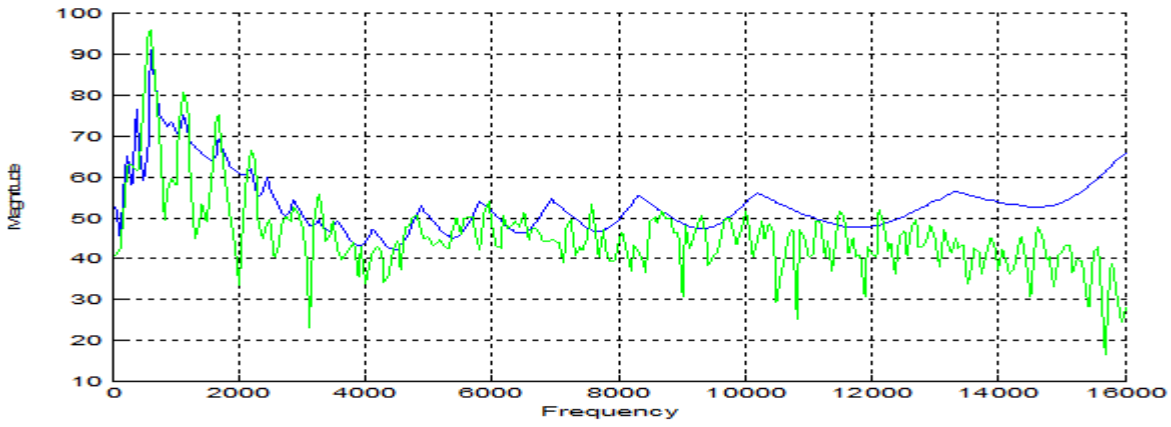
Nous avons déjà mentionné précédemment l'existence de bande critique ($ou\ Bark$) : le niveau de bruit masqué dépend uniquement de l'énergie du signal dans une bande critique, et non de sa répartition spectrale dans la bande.

Une analyse par bandes critiques compliquerait cependant de traitement du signal, de plus elle n'est pas absolument nécessaire. Plusieurs codes perceptuels découpent le spectre audio en bandes de largeurs égales.

Le filtrage en sous bande sera effectuée par une bande de filtre, le spectre complet du signal est découpé en intervalles de largeurs égales, ils correspondent à ce découpage de filtre de type passe-bande, dans un premier temps, la sortie de ces filtres est un signal (filtré) échantillonné à la fréquence originale fs.

Pour la première sous-bande (celle qui va de $L/N+1$), la décimation ou réduction de fs nemodifie pas le signal, puisque les fréquences utiliser sont inferieur à la nouvelle fréquence de Nyquist, mais pour les autre sous-bandes, le critère de Nyquist n'est plus vérifié : par conséquence il y à recouvrement.

Ce recouvrement va néanmoins s'opérer de manière ''propre'' de telle manière que le signal original puisse finalement être récupérer.



FigureIII.8 : Masque fréquentiel.

III.11 Pondération avec le masque temporel

Dans ce contexte, il reste à analyser le problème particulier de masquage temporel nous venons de voire que le niveau de bruit est constant dans une sous bande et dans un bloc. Il est déterminé par le nombre de bits attribués à la mantisse, est donc par ‘’hauteur ‘’ de seuil de masquage dans cette région de spectre.

III.12 Tatouage d’une trame

Le tatouage d’une trame dans un signal audio, c’est à dire la décision de classifier une trame en trame tatoué ou pas, peut passer par l’analyse trame à trame de l’énergie. Un changement soudain dans l’énergie ainsi qu’un changement soudain du contenu spectral peut s’apparenter à l’introduction d’un transitoire dans la trame, après cette technique en peut tatouer l’information inaudible à partir d’ajouter la signature à un partie de signal uniquement.

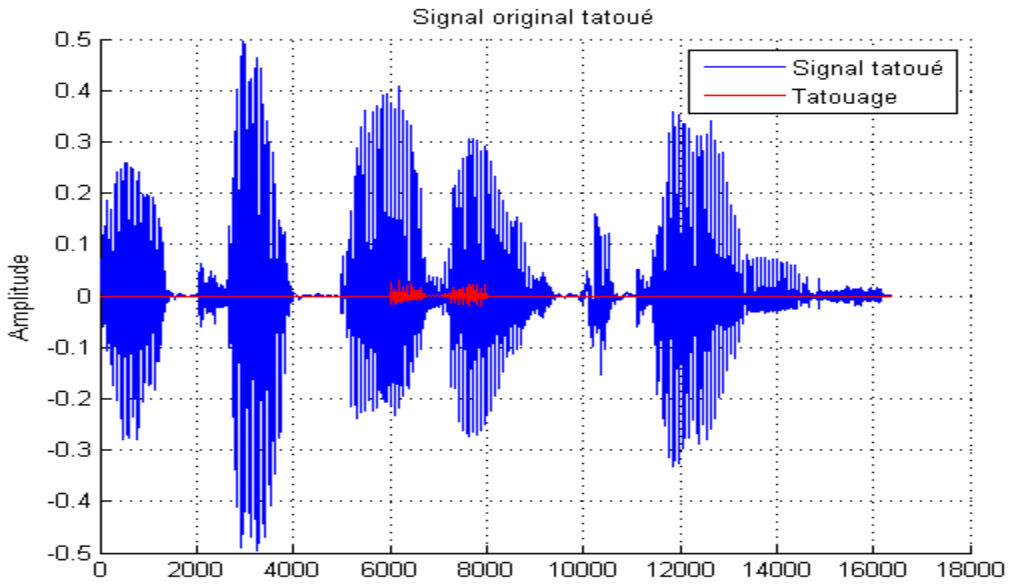
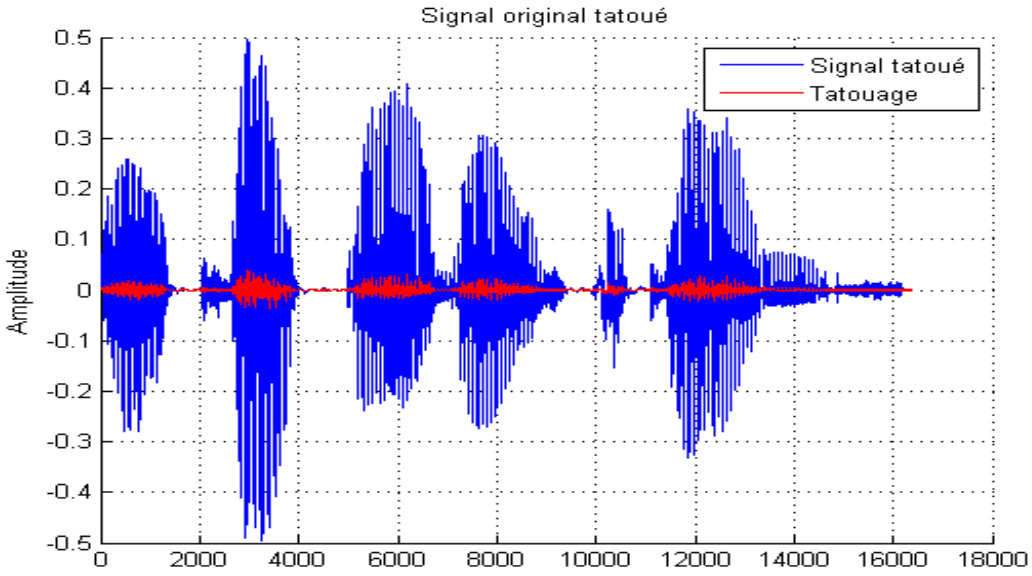


Figure III.9 : Tatouage d’une trame.

III.13 Tatouage global

Dans ce type de tatouage la séquence binaire d’ajouté par l’auteur ou signal audio transmis ou-bienla signature est sa pendant toute la période de signal, sans modifier ou toucher la forme du signal original.



FigureIII.10 : Signal original tatoué.

III.14 Extraction et vérification du tatouage

Garantir la vérification 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 réalises par un grand nombre d'auditeurs peut établir de manière fiable la transparence du tatouage.

Pour l'extraction, le signal audio original n'est pas nécessaire. Avant de commencer la détection, le signal est filtré par un filtre passe bas. Le signal obtenu est démodulé en le multipliant par le bruitutilisé lors de l'insertion de la signature. On calcule ensuite la corrélation entre la signatureet le signal obtenu. Et la détection de la signature est exécutée après le décodage entropique avec les mêmes procédures d'insertion sont employées pour extraire la signature.

III.15Fidélitéperceptuelle

Lapremièrecondition pournotresystèmedetatouage,estqu'enaucuncas,letatouagenedoit influersurl'informationcontenue danslesignal,ousurl'écoutede cedernier.C'estpourcelaque nous avons analysé l'adaptabilité du tatouage aux différents types de signaux susceptibles d'être tatoués.Eneffet,unsignalaudio peutcontenirdesinformationstrèsdifférentes lesunesdesautres, nous avons donc étudié l'effet du tatouage dans différentes parties du signal audio :

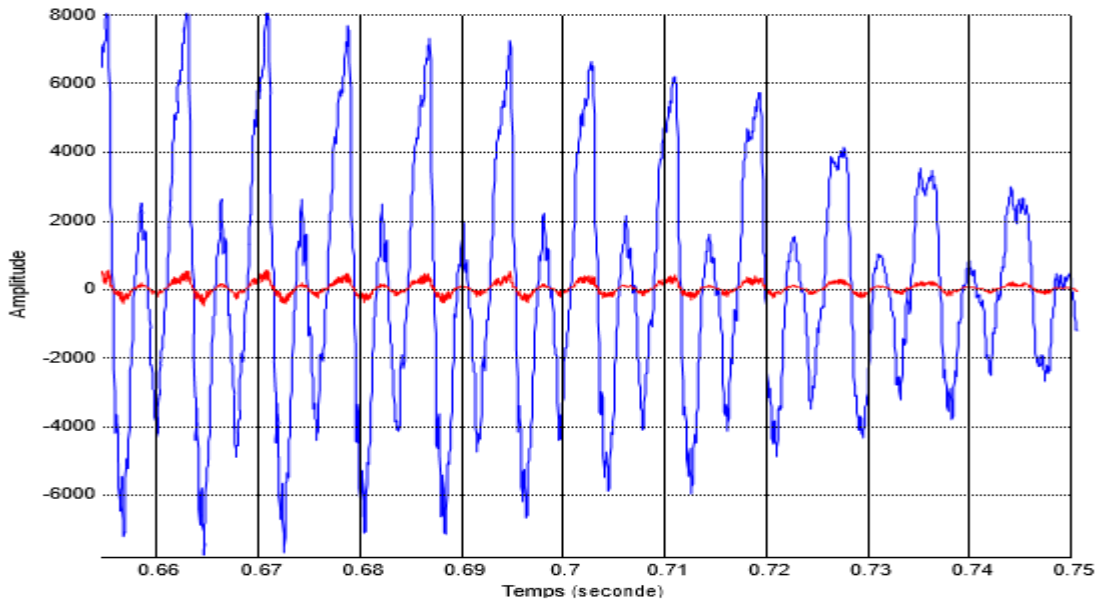
III.15.1 Partievoisée

Commedanslamajorité descas,lasignatureestreprésentéeparunbruit,lemasquagedecedernier devientdélicatquandilsetrouvedansunezone voiséedusignal.Eneffet,unbruitestplus difficilementmasquéparunsonvoisé,qu'un sonvoisé par un bruit. Rappelons à titred'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 laconfiguration inverse [6].

Celaexpliqueladiscriminationdessonstonalset non-tonalsdanslemodèlepsychoacoustiquedela normeMpeg1'niveau1'.Eneffet,ladéterminationdesseuilsdemasquagediffèreselonlanature du sonmasquant ; les formules déterminées empiriquement pour le calcul de la fonction d'étalement sont différentes selon que la composante soit tonale ou non.

La figure suivante montre une parcelle d'un tatouage représenté par un bruit unique, masquée par une partie voisée du signal tatoué :



FigureIII.11 : Partie voisée du signal tatoué.

Nous observons sur cette figure que le tatouage en rouge évolue en fonction du signal, on remarque aussi une certaine périodicité dans le tatouage qui est en harmonie avec la périodicité du signal.

III.15.2 Partie contenant du bruit

Il est plus aisé à un bruit de masquer une signature, mais en général, dans le cas des séquences audio destinées au tatouage, les sons non voisés (sons contenant du bruit) sont d'amplitude faible, cela implique une forte atténuation de l'amplitude du tatouage. Pour une bonne résistance aux perturbations, le tatouage doit de préférence avoir les amplitudes les plus élevées possibles, tout en tenant compte de la condition d'inaudibilité. Ce compromis est assez difficile à réaliser.

La figure suivant illustre une portion d'un tatouage masquée par une partie bruitée (son non voisé)du signal tatoué :

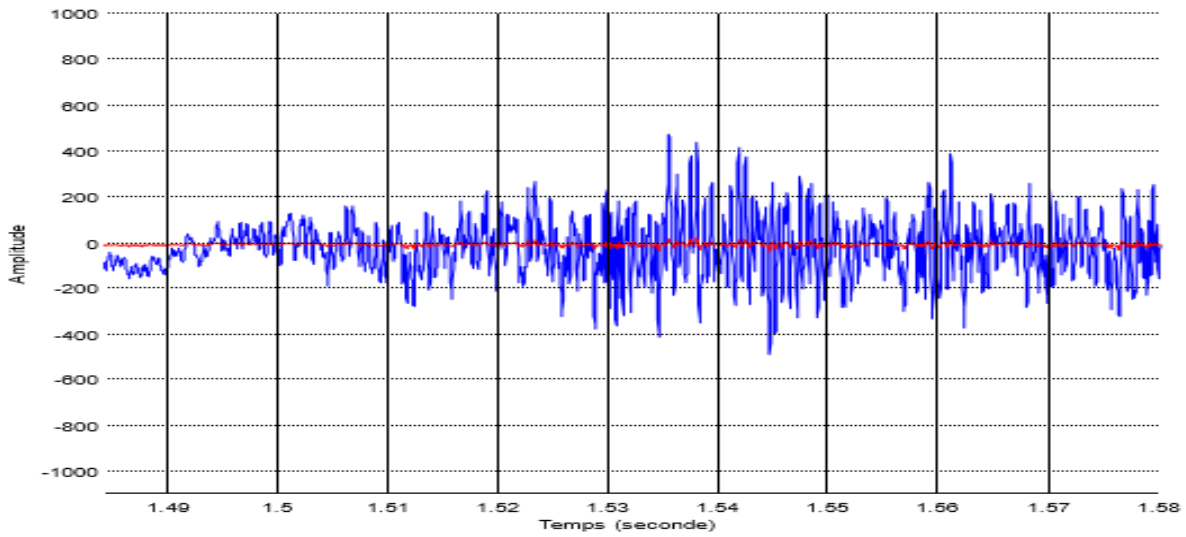
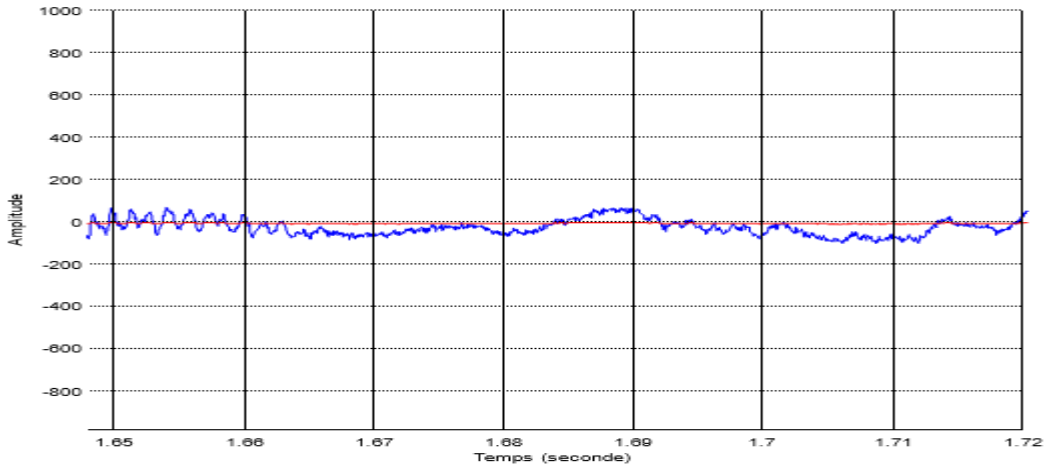


Figure III.12 : Partie bruitée du signal tatoué.

Nous observons que l'amplitude du tatouage est plus faible (indice≈40) comparativement aucasprécédent (indice ≈500), et que sa fréquence est plus élevée, suivant ainsi l'évolutiondes fréquences du signal.

III.15.3 Partie silencieuse

Il est évident que sur une partie de silence d'un signal audio le tatouage ne peut être masqué,car dans de une telle zone, l'amplitude du tatouage est quasiment nulle, comme illustré sur la figuresuivante :



FigureIII.13 : Partie silencieuse du signal tatoué.

Le tatouage sur cette figure a une amplitude quasiment nulle sur toute la parcelle silencieuse du signal.

A l'issue de ces observations, nous pouvons dire que notre tatouage s'adapte bien aux différentes natures de signaux que peut comporter une séquence audio ; cela lui permet d'être imperceptible dans tout les cas de figure. Cette propriété est due au double masquage fréquentiel – temporel.

En plus des observations précédentes, des tests d'écoutes sur différentes séquences (voix masculine, voix féminine, music…) de tailles différentes, nous permettent d'affirmer que l'inaudibilté du tatouage est assuré par notre algorithme.

III.16 Fiabilité des résultats

Pour prouver la propriété intellectuelle d'un média, la méthode utilisée doit être crédible et fiable. Le test de similitude de notre algorithme $\sin(ew, w)$peut être considéré comme un outil crédiblede jugement objectif.

III.16.1Cas sans transmission (local)

Après plusieurs tests, dans le cas idéal, c'est-à-dire qu'aucune perturbation n'est subie par le signal tatoué (sans transmission), notre test de similitude retourne le résultat :

$\sin(ew, w) = t$, c'est-à-dire 100% de similitude entre le tatouage extrait et le tatouage d'origine.

III.16.2 Cas d'une transmission

Dans le cas d'une transmission, des perturbations sur le signal tatoué doivent être prises en compte. Les résultats de similitude peuvent différer selon la chaine d'opération que subit le signal(compression, filtrage, codage, transmission…). Un seuil de tolérance : $s = \sin(ew, w)$doit être fixé selon le cas de transmission et selon les besoins, en veillant à ce que le seuil doit être leplus proche possible de 1. Ce point n'a pas été testé en détail dans notre travail, nous le suggérons en perspective de travaux à venir.

III.17 Robustesse

Le tatouage doit être robuste aux attaques licites (compression, transmission…), comme aux attaques illicites (piraterie).

Dans le cas de la signature représentée par un bruit, le tatouage est statistiquement indétectable sans connaissances à priori. Concernant la signature représentée par une séquence audio propre à l'auteur, la détection de celle-ci par un pirate n'est pas impossible, mais reste difficile. De plus, le pirate n'aura atteint son but qu'en modifiant ou supprimant le tatouage, ce qui risque d'altérer le signal hôte et le rendre inexploitable.

III.18Résultats de la stéganographie

III.18.1 Système QIM-MS-VQ appliqué au quantificateur des ISFs

Pour mettre en œuvre ce système, que nous avons appelé QIM-MS-VQ, nous nous sommes intéressés à deux techniques de modification (c.f, chapitre 3). Il s'agit de la procédure de dissimilation par remplacement avec tramage (DRT) et la procédure classique de dissimulation par remplacement simple (DRS). Rappelons qu'une partie du débit du deuxième étage du S-MSVQ des ISFs sera allouée pour dissimuler les messages secrets binaires. D'une façon similaire au système précédent, il est toujours important d'évaluer les performances de ce quantificateur pour différents choix de dictionnaires utilisés afin de retenir le choix optimal pour chaque débit de dissimulation.

Dans les simulations suivantes, nous utilisons la même base de données des ISFs qui contient 180000 vecteurs. Les vecteurs ISFs sont quantifiés par un quantificateur S-MSVQ de G.722.2 modifié selon le principe de la technique QIM-MS-VQ.

III.18.1.1 Performances des systèmes stéganographiques QIM-MS-VQ

Les performances des systèmes stéganographiques QIM-MS-VQ, conçus selon les deuxtechniques de modification (DRS et DRT) sont présentées sur les tableaux 4.7 et 4.8. Pour un débit de dissimilation fixe, les dictionnaires notés dans les tableaux sont seulement les dictionnaires du deuxième étage (dic21,…, dic25) du S-MSVQ. Précisons que le "1" signifie que le dictionnaire correspondant est utilisé dans la procédure de dissimulation. Le débit de ce dictionnaire est alors réservé pour le message à dissimuler.

Débit de dissimulation (Bits/trame)	**SD Moy. [dB]**	**SD "Outliers" (en %)**	
		2-4 dB	**> 4 dB**
5(0-0-0-0-1)	1.65	21.63	0.08
6 (1-0-0-0-0)	2.34	60.79	3.20
7 (0-1-0-0-0)	1.92	39.39	1.11
10(0-0-0-1-1)	2.20	58.38	0.78
11(1-0-0-0-1)	2.72	75.34	6.40
12(0-0-1-0-1)	2.39	65.97	2.53
14(0-1-1-0-0)	2.61	71.19	5.66
16(1-0-0-1-1)	3.09	80.70	12.65
17(0-1-0-1-1)	2.81	81.93	6.35
18(1-0-1-0-1)	3.27	76.25	18.82

20(1-1-1-0-0)	3.41	71.99	24.02
23(1-1-0-1-1)	3.57	69.77	28.64
24(0-1-1-1-1)	3.29	79.05	17.93
25(1-1-1-0-1)	3.67	65.67	33.00
30(1-1-1-1-1)	3.98	53.12	46.57

Tableau III.1 : Performance du système QIM-MS-VQ conçu par la méthode DRS.

Débit de dissimulation (Bits/trame)	**SD Moy. [dB]**	**SD "Outliers" (en %)**	
		2-4 dB	**> 4 dB**
5(0-0-0-0-1)	1.48	14.82	0.06
6 (1-0-0-0-0)	2.03	46.36	1.64
7 (0-0-1-0-0)	1.79	31.98	0.64
10(0-0-0-1-1)	1.92	39.05	0.40
11(1-0-0-0-1)	2.35	62.14	2.70
12(0-0-1-0-1)	2.13	51.38	1.37
14(0-1-1-0-0)	2.40	64.21	3.31
16(1-0-0-1-1)	2.67	73.41	5.93
17(0-0-1-1-1)	2.48	69.77	3.20
18(1-0-1-0-1)	2.84	75.79	9.05
20(1-1-1-0-0)	3.10	76.91	14.50
23(1-0-1-1-1)	3.12	77.40	14.95
24(0-1-1-1-1)	2.95	80.70	9.85
25(1-1-1-0-1)	3.29	76.03	19.30
30(1-1-1-1-1)	3.54	69.37	28.53

Tableau III.2 : Performance du système QIM-MS-VQ conçu par la méthode DRT.

La quantification se fait à des débits de dissimulation variant de 5 à 30 bits/trame où l'onprésentele choix optimum des dictionnaires utilisés dans la dissimulation de message selon la distancespectrale. Les messages à dissimuler sont générés d'une façon aléatoire.

Ces résultats montrent que la dégradation des performances SD due à la dissimulation n’est pas proportionnelle au débit de dissimulation. Par exemple, pour un débit de dissimulation de 10bits/trame, la dégradation causée par la dissimulation dans les indices des dictionnaires dic24

etdic25 est inferieure à celle causée par la dissimulation dans les indices du dictionnaire dic21 du cas de 6 bits/trame. En effet, la dégradation est plutôt liée à l'importance du dictionnaire utilisé. Sachant que la perception humaine de la distorsion causée par les formants dans les bandes de fréquence élevées est négligeable par rapport à celle dans les basses fréquences, donc les dictionnaires qui représentent les hautes fréquences sont moins importants que ceux des basses fréquences. On remarque aussi que les deux techniques ont le même comportement en ce qui concerne l'importance des dictionnaires.

Pour tous les débits de dissimulation, l'évaluation comparative des résultats montre que le système stéganographique QIM-MS-VQ conçu par la méthode DRT assure de meilleures performances que le QIM-MS-VQ conçu pas la technique DRS. D'autre part, on constate que pour les deux systèmes stéganographiques, la quantification de qualité transparente n'est pas totalement assurée même pour un débit de dissimulation minimal de 5 bits/trame en utilisant le dictionnaire le moins important(dic25).

III.18.1.2 Performance du G.722.2 avec implémentation du QIM-MS-VQ des ISFs

Pour les simulations suivantes, nous utilisons toujours le G.722.2 en mode 12.65 kbps où 46 bits/trame sont utilisés pour coder les paramètres ISF. La base de données utilisée est composée de 10 séquences de parole de 32 secondes extraites de la base de données TIMIT.

Les performances du codeur G.722.2, avec implémentation du système QIM-MS-VQ au niveau du quantificateur S-MSVQ des ISFs, sont présentées sur le tableau 4.9. Pour chaque méthode de dissimulation utilisée (DRS ou DRT), les résultats sont donnés en termes de SDmoyenne pour l'évaluation du codage des ISFs avec dissimulation et en notes WB-PESQ pour l'évaluation de la qualité globale de la parole synthétisée par le G.722.2. Notons que pour chaque débit de dissimulation, les choix optimaux des dictionnaires du 2ème étage utilisés dans la procédure de dissimulation sont les mêmes que ceux trouvés précédemment. Enfin, le message binaire secret dissimulé est le flux binaire généré (bit-stream) par le codeur de parole MELP de 2.4 kbps en utilisant la même base de données arabe précédente. Rappelons que les performances données pour un débit de dissimulation 0 bits/trame sont exactement les performances du codeur standard original G.722.2 sans modification.

Ces résultats montrent que l'imperceptibilité n'est pas liée seulement à la capacité (débit) de dissimulation mais aussi au choix des dictionnaires utilisés. Par exemple, pour les débits de dissimulation de 23 et 24 bits/trame (1150 et 1200 bps), on remarque que l'imperceptibilité de la dernière est meilleure. Donc dans ce cas, l'utilisation de 24 bits /trame offre plus de capacité et une meilleure imperceptibilité.

On remarque aussi que l'imperceptibilité de la technique QIM-MS-VQ conçu par DRT est meilleure que celle conçu par DRS pour tous les débits de dissimulation.

Débit de dissimulation (Bits/trame)	QIM-MS-VQ par DRS		QIM-MS-VQ par DRT	
	SD Moy.	WB-PESQ	SD Moy.	WBPESQ
0	0.91	3,790	0.91	3.790
5	1.58	3.775	1.44	3.738
6	2.06	3.265	1.88	3.321
7	2.08	3.684	1.91	3.728
10	2.09	3.651	1.86	3.651
11	2.51	3.233	2.22	3.337
12	2.52	3.665	2.22	3.700
14	2.89	3.556	2.57	3.568
16	2.86	3.210	2.50	3.279
17	2.89	3.579	2.52	3.574
18	3.10	3.224	2.75	3.312
20	3.54	3.054	3.15	3.110
23	3.51	3.093	3.07	3.161
24	3.52	3.480	3.07	3.553
25	3.77	3.072	3.34	3.105
30	4.02	3.033	3.52	3.139

Tableau III.3 : Performance du codeur G.722.2 avec implémentation du système QIM-MS-VQ des ISFs.

Nous donnons ci-dessous, un exemple de résultat de simulation obtenu lors du codage de la même phrase utilisée dans la partie précédente (Shehadyourdark suit in greasywash water all year). Les performances du codeur G.722.2 (original et avec dissimulation du message) sont données en termes de WB-PESQ sur les figures suivantes, ou l'on représente aussi les différents formes d'onde synthétiques ainsi que l'évolution desSD correspondants pour plusieurs débit de dissimulation.

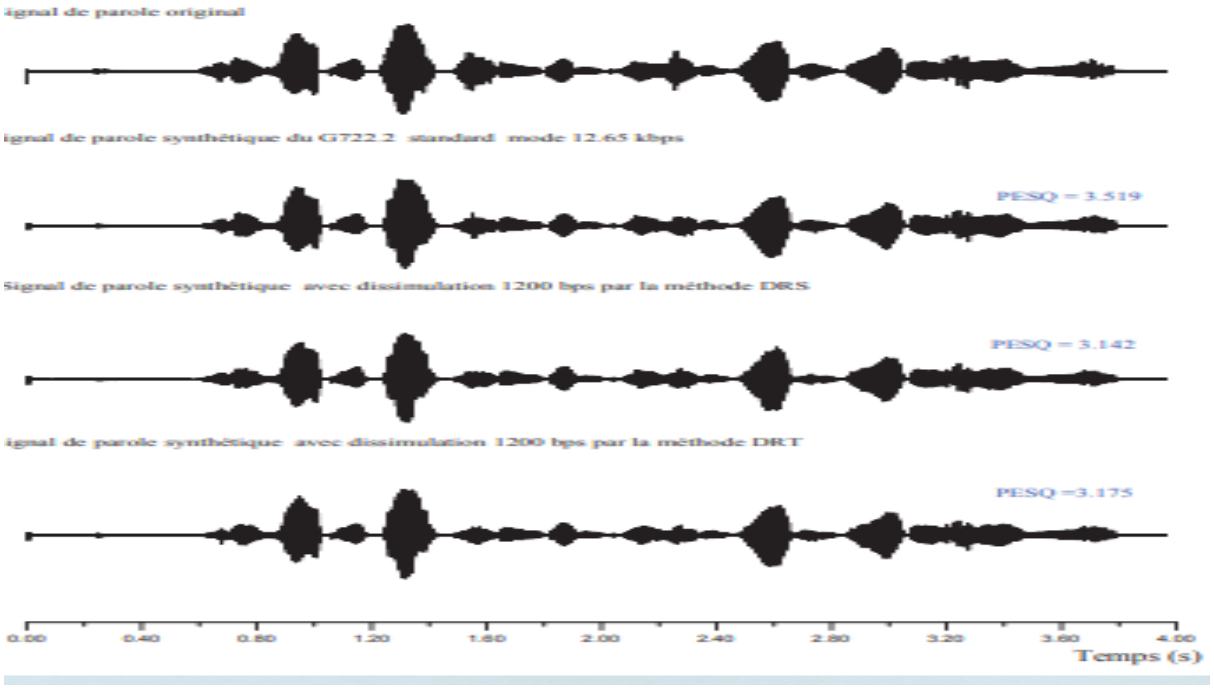


Figure III.14 : Formes d'ondes synthétiques avec dissimulation QIM-MS-VQ par DRS et DRT.

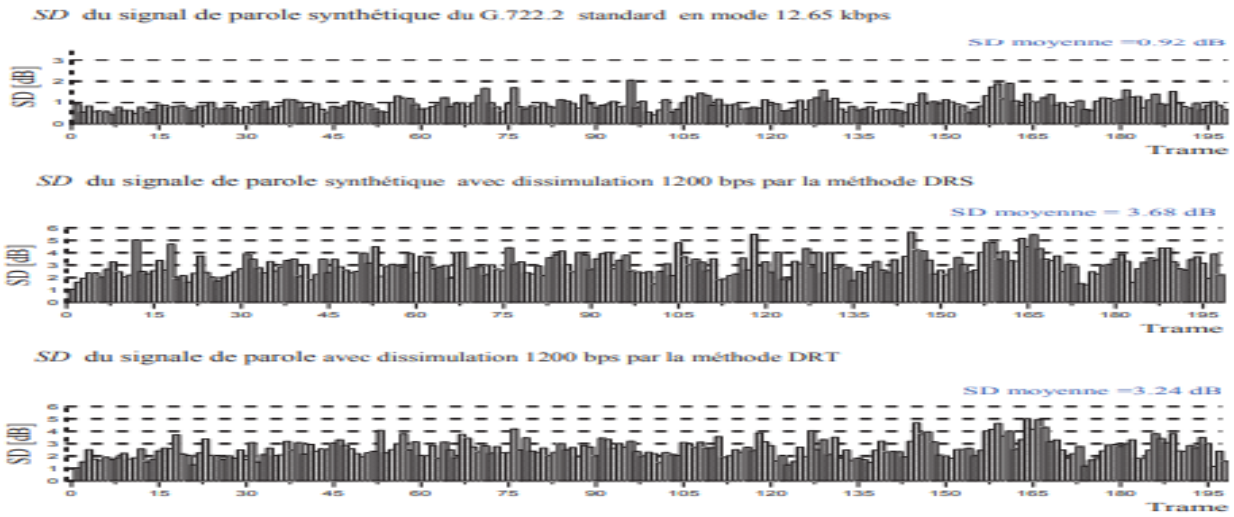


Figure III.15 : Comparaison de l'évolution de la SD entre DRS et DRT.

III.18.2La stéganographie a l'excitation du G.722.2

Dans cette méthode, on utilise l'indice du vecteur d'excitation (dictionnaire algébrique) pour dissimuler le message secret. On modifie donc la boucle d'analyse par synthèse du codeur de parole standard AMR-WB G.722.2. Rappelons que le vecteur d'excitation est calculé chaque sous-trame de 5 ms en utilisant une méthode d'entrelacement (c.f. chapitre 2). Chaque soustrame est représentée par 4 pistes où un nombre maximum de 4 bits peut être dissimulé par piste. On a donc une possibilité de dissimuler jusqu'à 16 bits par sous trame comme on l'a bien expliqué dans le

chapitre précédent. Le problème majeur dans cette méthode est de trouver la distribution optimale de ces bits sur les 4 pistes afin d'assurer une dégradation minimale.

Pour un même débit de dissimulation, nous présentons ci-dessous les performances en termes de WB-PESQ des meilleures distributions des bits/sous-trame que nous avons trouvé.

La base de données parole utilisée dans les simulations suivantes se compose de 100 phrases de 347 secondes, prises de la base de données TIMIT.

Pour des débits de dissimulation supérieurs à 2 bits par sous-trame, les WB-PESQ moyens obtenus sont légèrement inferieurs au WB-PESQ moyen obtenu par le codeur standard G.722.2 sans dissimulation. Cependant, ces performances restent très acceptables justifiant ainsi l'imperceptibilité de la technique de dissimulation implémentée.

En plus de la dissimulation des messages secrets, cette technique a permit aussi de diminuer la complexité de codage en diminuant le nombre de combinaisons possibles nbcpour générer le vecteur d'excitation. Pour un débit de 16 bits par sous-trame qui est équivalant à 3200 bps le nombre nbc est égal à 256. En le comparant avec le nombre nbc du codeur standard (1280) onconclu que la complexité dans cet étage est diminuée de 20% compareé à celle du codeur standard.

Débitde dissimulation (Bits/sous-trame)	**performance du G.722.2 : WB-PESQ moyen**
1 (0+0+1+0)	3.832
2 (0+0+1+1)	3.809
3 (1+1+0+1)	3.797
4 (0+1+1+2)	3.777
5 (3+0+1+1)	3.756
6 (0+2+1+3)	3.740
7 (1+3+3+0)	3.709
8 (2+1+4+1)	3.672
9 (2+1+2+4)	3.659
10 (3+4+2+1)	3.624
11 (2+4+3+2)	3.594
12 (2+3+3+4)	3.554
13 (3+4+4+2)	3.502
14 (2+4+4+4)	3.455
15 (4+3+4+4)	3.414
16 (4+4+4+4)	3.353

Tableau III.4 : Performance du G.722.2 avec dissimulation dans l'indice de l'excitation.

Nous présentons, dans la figure 4.5, l'évolution du nombre nbc en fonction des débitsdedissimulation. Le nombre nbc est calculé par la distribution optimale des bits obtenu précédemment. On remarque que le nombre nbc est supérieur au nombre nbc du standard seulement pour les deux débits de dissimulation de 1 et 2 bits/sous-trame. Donc la complexité est plus petite que la complexité du codeur standard G.722.2 pour chaque débit supérieur à 2 bit/soustrame (400 bps).

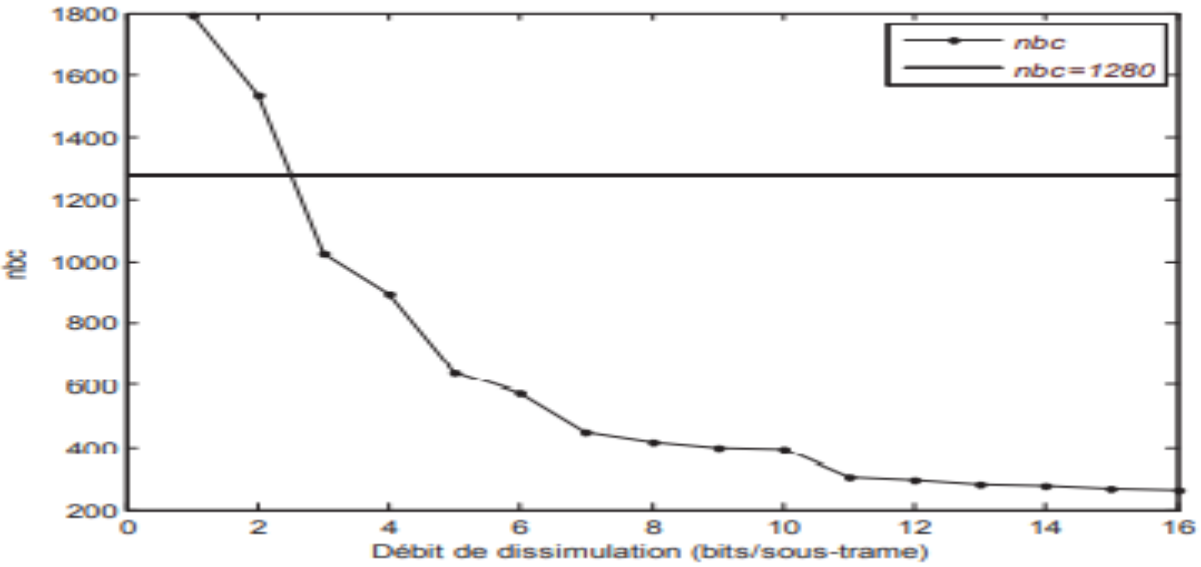


Figure III.16 : Complexité de la QIM-DD appliquée sur l'indice d'excitation**.**

III.18.3 Dissimulation dans le délai tonal

Cette méthode de dissimulation est appliquée au niveau de la quantification du délai tonal (pitch) du G.722.2. Il s'agit de la méthode LSB adaptative qui modifie la valeur du délai tonal en fonction du message à dissimuler (c.f. chapitre3). Rappelons que pour chaque trame le délai tonal est calculé deux fois dans les sous-trames 1 et 3. Il est quantifié sur neuf bits, donc la valeur quantifiée varie dans l'intervalle [0,…, 512] dans le mode 12.65 kbps. Par la suite, on utilisera plusieurs combinaisons du nombre de bits à dissimuler et des seuils pour étudier l'influence de ces paramètres sur la qualité de la parole publique.

Les performances du codeur G.722.2, avec implémentation de la technique de dissimulation dans le délai tonal sont données sur le tableau 4.11. Nous utilisons la même base de données de parole de 100 phrases. Les performances sont données en termes de WBPESQ en fonction du nombre de bits dissimulés n et des seuils prédéfinis. Le message secret est généré d'une façon aléatoire.

(Seuil, n)	Performances G.722.2 : WB-PESQ moyen
(64, 1)	3.768
(64, 2)	3.595
(64, 3)	2.916
(64, 4)	2.199
(64, 5)	1.706
(128, 1)	3.776
(128, 2)	3.633
(128, 3)	3.077
(128, 4)	2.443
(128, 5)	1.917
(256,1)	3.789
(256, 2)	3.680
(256, 3)	3.278
(256, 4)	2.828
(256, 5)	2.467
(384, 1)	3.779
(384, 2)	3.713
(384, 3)	3.555
(384, 4)	3.381
(384, 5)	3.181

Tableau III.5 : Performance du G.722.2 avec dissimulation dans le délai tonal.

D'après ces résultats, on remarque que l'imperceptibilité est liée au nombre de bits modifié dans chaque indice présenté dans le tableau par n et au seuil de comparaison. Les meilleurs résultats sont obtenus lorsqu'on utilise un grand seuil et un nombre n petit. Par exemple, l'utilisation d'une paire (seuil, n) égale à (384,1) donne une note WB-PESQ moyen égale à 3.779 qui représente le meilleur résultat possible dans notre cas.

D'autre part, les deux paramètres seuil et n ont une grande influence sur la capacité moyenne de dissimulation (débit). Dans ce qui suit, nous donnons les résultats de capacité obtenus pour les différentes paires (seuil, n).

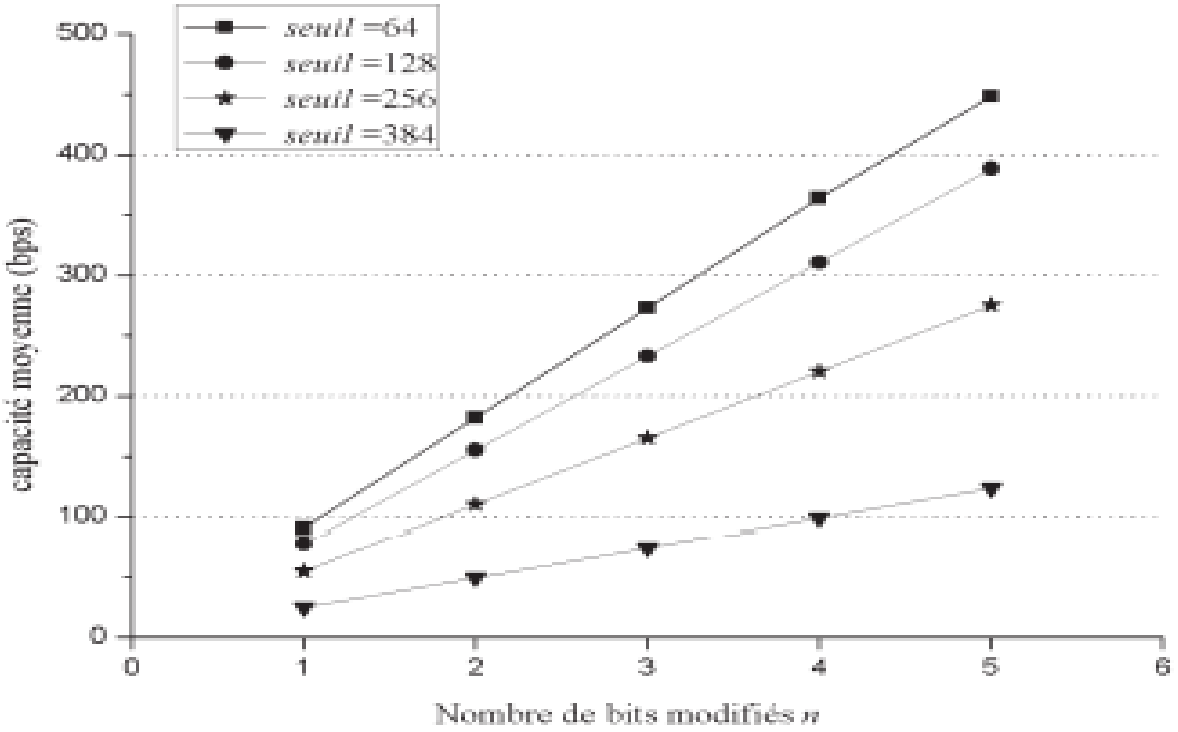


Figure III.17 : Capacité de la méthode LSB adaptative.

La figure 4.6 montre que l'utilisation d'un petit seuil offre un débit de dissimulation plus grand pour chaque nombre n. Par exemple, l'utilisation de seuils égaux à 64 et 384 avec un nombre n = 4 donne des débits de dissimulation de 365 bps et de 98 bps respectivement.

On remarque que l'utilisation d'une paire (384, 4) donne une capacité moyenne de dissimulation très proche à celle donnée par une paire (64, 1)

D'après l'étude de la capacité moyenne et l'imperceptibilité, onconclu que ces performances dépendent aux paramètres (seuil, n) sélectionnés. Néanmoins, le plus important dans notre système stéganographique est la relation entre la capacité moyenne et l'imperceptibilité. On remarque que l'utilisation d'un seuil = 64 donne les meilleurs performances (capacité moyenne / imperceptibilité). Par exemple si on prend les deux paires (64, 1) et (384, 4), on remarque qu'ils donnent des capacités moyennes de dissimulation très proches (91 bps et 99 bps) avec des WB-PESQ moyens égals à 3.772 et 3.347respectivement. Donc l'imperceptibilité de (64, 1) est supérieure avec la même capacité Pratiquement.

III.19Conclusion :

Le tatouage et la stégqnogrqphiepeut être utilise donc pour améliorer les systèmes de transmissions existants. En effet, il peut être utilise pour :

-Aider à récupérer le signal hôte transmis sur le canal de communication par la correction des erreurs, l'information cachée est utilisée pour réduire le bruit lie à la transmission.

-Tester la qualité de services des transmissions multimédia

-Evaluer, de manière subjective, la qualité du signal de parole transmis sur la bande téléphonique.

-Elargir la bande passante du signal transmis dans un système de communication.

Conclusion Générale

Cetprojet traite du tatouage et stéganographie des signaux audio numériques dans le contexte suivant protection des droits d'auteur. Cette classe d'application permet d'identifier l'auteur ou lepropriétaire légal d'une œuvre audionumérique.

Le but principal de ce travail est de concevoir un tel système de tatouage ou de stéganographie. Ces conceptionssontsoumise à la principale contrainte d'inaudibilité, qui est contournée grâce au masquage fréquentiel-temporel de notre procédures de tatouage et stéganographie.

Le langage de programmation utilisé dans ce présent travail est MATLAB.

Référencesbibliographiques

[1] Ingemar J. Cox: Digital watermarking and steganography. Morgan Kaufmann, Burlington, MA, USA, 2008

[2] Frank Y. Shih: Digital watermarking and steganography: fundamentals and techniques. Taylor & Francis, Boca Raton, FL, USA, 2008

[3] Xiang, Y.; Hua, G.; Yan, B. Digital Audio Watermarking: Fundamentals, Techniques and Challenges; Springer: Singapore, 2017.

[4] I.J. Cox, M.L. Miller, and A.L. Mc Kellips. " Watermarking as communications with side information". Proceedings of the IEEE, Special Issue " Identification and protection of multimedia information". July 1999.

[5] V. Martin. Contribution Des Filtres LPTV Et Des Techniques D'interpolation Au Tatouage Numérique. Thèse de doctorat. L'institut National Polytechnique De Toulouse, 2006.

[6] HAICHEUR. Tatouage par DWT-QR Application à l'Indexation des Documents Audio-visuels, 2019.

[7] M.Parvaix. Séparation De Sources Audio Informée Par Tatouage Pour Mélanges Linéaires Instantanés Stationnaires. Thèse de doctorat, Université De Grenoble Institut Polytechnique De Grenoble, 2010.

[8] DALIL BATTIKH: ''Sécurité de l'information par stéganographie basée sur les séquences chaotiques''. INSA de Rennes, 2015.

[9] 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, january2003.

[10] 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, October2000.

[11] 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, january2002.

[12] Ingemar Cox, Matthew Miller et Jeffrey Bloom: Digital watermarking. Morgan Kaufmann Publishers, San Francisco, USA, 2002.

[13] John F. Tilki et A. A. Beex: Encoding a hidden digital signature onto an audio signal using psychoacoustic masking. Dans 7th International Conférence on Digital Signal Processing Applications and Technology, pages 476–480, Boston, MA, USA, october 1996.

[14] CLEO BARAS: ‘’Tatouage informé de signaux audio numériques’’, thèse de doctorat Ecole Nationale Supérieure des Télécommunications, (paris), 2005.

[15] Philips: Content identification. http: //www.research.philips.com/initiatives/ contented/index.html

[16] Teletrax: Global television tracking and reporting. http://www.teletrax.tv

[17] 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, october2001.

[18] 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.

[19] SOFIANE BRACI: ‘’Etude des méthodes de dissimulation informées de données appliquées aux supports multimédias’’. Thèse de doctorat, Université Paris Sud - Paris XI, 2010.

[20] DELENDA Sabah : Méthodes pour la dissimulation d’information dans une image, thèse de doctorat Université Mustapha Ben Boulaid Batna2 Faculté des Mathématiques et d’informatique, Batna, 2019.

[21] S. Katzenbeisser, F.A.P. Petitcolas, Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, Norwood, MA, 2000.

[22] Sridevi R., Damodaram A., SVL.Narasimham, Efficient Method of Audio Steganography by Modified LSB Algorithm and Strong Encryption Key with Enhanced Security, Journal of Theoretical and Applied Information Technology, 2009.

[23] G. J. Simmons, "The prisoners' problem and the subliminal channel" in Proc. Advances in Cryptology (CRYPTO '83), pp. 51-67.Berglund, J.F. and K.H. Hofmann, 1967.Compact

semitopological semigroups and weakly almost periodic functions. Lecture Notes in Mathematics, No. 42, Springer-Verlag, Berlin-New York.

[24] Katzenbeisser S. &Petitcolas F. Information Hiding Techniques for Steganography and Digital Watermarking. Artech House Inc, 2000.

[25] Dunbar B. “A detailed look at Steganographic Techniques and their use in an Open Systems Environment”, SANS Institute, 2002.

[26] Sellars D.“An Introduction to Steganography”, University of Camberge, 2003.

[27] Yang Y. “Digital Watermarking Technology”, Faculty of Computer Science, Dalhousie University, URL: http://www.cs.dal.ca/~yyang/6505/6605.pdf, 2001.

[28] Polpitiya A.D., & Khan W.J. Information Hiding in Audio Files with Encryption”, Data Security Project, Washington University, 2001.

[29] Bender W., Gruhle D., Morimoto N., & Lu A. Techniques for Data Hiding”, IBM System Journal, vol.35, [online]: available at: http://isj.www.media.mit.edu/isj/SectionA/ 313.pdf, 1996.

[30] RobertKrenn, “Steganography and steganalysis”, An Article, January 2004.

[31] NedeljkoCvejic, TapioSeppben “Increasing the capacity of LSB-based audio steganography ” FIN90014 University of Oulu, Finland ,2002.

[32] W. Bender, W. Butera, D. Gruhl, R. Hwang, F. J. Paiz, S. Pogreb, “Techniques for data hiding”, IBM Systems Journal, Volume 39 , Issue 3-4, July 2000, pp. 547 – 568

[33] F.M.M. Bouyé, « Application des codes correcteurs d’erreurs en stéganographie », Thèse, Faculté des Sciences, Univ. Mohammed V Agdal, Rabat, 11 juillet 2012.

[34] W. Bender, D Gruhl, and N. Morimoto, “Technique for data hiding”, Proc, of the SPIE, 1995.

[35] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, “Secure Spread Spectrum watermarking for Multimedia”. Tech. Rep. 95-10, NEC Research Institute, 1995.

[36] E. Zwicker, R. Feldtkeller : « Psychoacoustique L’oreille récepteur d’information », Stuttgart, West Germany : HirzelVerlag, 1967 (Masson 1981, traduit de l’allemand par C.SORIN).

[37] T. Painter, A. Spanias : « Perceptual Coding of Digital Audio », Travaux de l’IEEE, Vol.88.No.4, 2000.

[38] Calliope : « La parole et son Traitement Automatique », Edition : Masson, 1989.

[39] L. Gomes : « Tatouage de signaux audio ». Thèse de doctorat, Université Paris V, 2002.

[40] R. Garcia : « Digital watermarking of audio signals using psychoacoustic auditory model and spread spectrum theory », in '107th Convention of Audio Engineering Society (AES)', New York, USA, 1999.

[41] Norme internationale ISO/CEI 11172-3 (MPEG1, partie : Audio) : « Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1.5 Mbits/s », 1993.

[42] P. Noll : « Wideband speech and audio coding», IEEE Communications, pp. 34-44,1993.

[43] lSO, http://www.iso.org, Site officiel de l'ISO (International Organization for Standardization).

[44] A. Le Guyarder, P. Philippe, J. B. Rault : « Synthèse des normes de codage de la parole et du son (UIT-T, ETSI ET ISO/MPEG) », pp. 425-441. Ann. Télécommunication, 55, n° 9-10, 2000.

[45] É. Fert, S. Jeannin : « Compressions MPEG-1 à MPEG-4 », Division traitement numérique du signal aux laboratoires d'électronique Philips, Techniques de l'Ingénieur, traité Télécoms.

[46] http://www.fraunhofer. de, Site officiel de La 'Fraunhofer-Gesellschaft', organisme allemand dédié à la recherche en sciences appliquées.

[47] M.Barni, F. Batrolini: « Watermarking systems engineering », Edition: Marcel Dekker, 2004.

[48] D. Swanson, B. Zhu, H. Tewfik, and L. Boney, « Robust audio watermarking using perceptual masking », Department of Electrical Engineering, Université de Minnesota, 1996.

[49] M. Auvrey : « Traitement du signal », IST Jussieu Paris 6.

[50] Frank R.Kshischang : « The Hilbert Transform », The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, Université de Toronto, 2006.

[51] http://research.microsoft.com/apps/search/default.aspx?q=MSR+Identity

[52] PH.GASSER: ''Les formats MPEG audio'', MSH Paris Nord-plats-forme Arts, sciences, Technologie. (Décembre 2006).

[53] ITAN: ‘’La compression des signaux audio et vidéo’’.2010.

[54] E.Zwiker, R.Feldtkelle:‘’Psycho-acoustique L’oreille récepteur d’information’’, Stuttgart, West Germany: HirZelVerlag, 1967 (Masson 19981, traduit de l’allemand par C.SORIN).

[55] SOFIANE BRACI: ‘’Etude des méthodes de dissimulation informées de données appliquées aux supports multimédias’’. Thèse de doctorat, Université Paris Sud - Paris XI, 2010.

[56] DOMINIQE, GIBERT : ‘’Eléments de Traitement du signal’’, (p7, p134), Août 1994

Annexe

ملخص

يوفر ظهور الملفات الرقمية إمكانيات مبتكرة في مجال تكنولوجيا البيانات ، يمكن للملفات المضيفة ، في حالتنا ، إخفاء المعلومات الرقمية. تسمى تقنيات الدفن هذه بالوشم وإخفاء المعلومات. عملنا على إعداد إجراء خاص بالعلامات المائية وإخفاء المعلومات يسمح بدمج حماية حقوق التأليف والنشر لبيانات الصوت الرقمي من خلال تعديل عينات الصوت مباشرةً. تستخدم هذه العملية بشكل مباشر إخفاء الإدراك والوقت والتردد للتأكد من أن وضع العلامات المائية الرقمية وإخفاء المعلومات غير مسموع وقوي.

الكلمات المفتاحية:وشم صوتي ، إخفاء صوتي ، علم صوتي نفسي ، إخفاء إدراكي.

Résumé

L'émergence des fichiers numériques offre des possibilités innovantes dans le domaine de la technologie des données, les fichiers hôtes, dans notre cas, les signaux audio, peuvent masquer des informations numériques. Ces techniques d'enterrement s'appellées le tatouage et la stéganographie. Notre travail à mettre en place une procédurede tatouage et de stéganographie qui permet l'intégration de la protection du droit d'auteur pour des données audio numériques en modifiant directement les échantillons audio. Ce processus utilise directement le masquage de perception, de temps et de fréquence pour garantir que le tatouage et la stéganographie numériques sont inaudibles et robustes.

Mots clés : Tatouage audio, stéganographie audio, psycho-acoustique, Masquage perceptuel.

Abstract

The emergence of digital files offers innovative possibilities in the field of data technology, host files, in our case audio signals, can mask digital information. These burial techniques are called tattooing and steganography. Our work to set up a watermarking and steganography procedure that allows the integration of copyright protection for digital audio data by directly modifying the audio samples. This process directly uses perception, time and frequency masking to ensure that digital watermarking and steganography is inaudible and robust.

Keywords:. Audio tattooing, audio steganography, psycho-acoustics, Perceptual masking.