

# Tatouage audio numérique robuste et haute capacité dans le domaine spectral basé sur des mesures perceptuelles

M. Bellaaj et K.Ouni

Unité de Recherche Systèmes Mécatroniques et Signaux  
Ecole Supérieure de Technologie et d'Informatique  
Université de Carthage  
Maha\_bellaaj@yahoo.fr  
Kais.ouni@esti.rnu.tn

## Résumé

*Dans cet article, nous allons présenter une technique de tatouage audio numérique opérant dans le domaine spectral. Le mappage temps-fréquence est réalisé à l'aide de la transformation MDCT [1] (Modified Discrete Cosine Transform). Pour chercher les positions d'insertion les moins audibles et sensibles à l'oreille humaine, nous allons présenter deux variantes de cette technique basées sur une modélisation du système auditif. Dans la première variante, nous allons exploiter les propriétés du modèle psychoacoustique 1 (MPH1) du codeur MPEG 1[2] couche 1. Dans la deuxième variante, nous allons intégrer le modèle psychoacoustique 2 (MPH2) du codeur MPEG 1 couche 3. Pour garantir plus d'inaudibilité, nous allons insérer les bits de la marque dans l'LSB (Least Significant Bits) des composantes déjà cherchées. Dans les deux variantes de la technique nous avons dupliqué les bits de la marque dans le but d'avoir une capacité d'insertion maximale et une meilleure robustesse contre les attaques. De plus, un code correcteur d'erreur (codage Hamming) a été utilisé pour rendre la détection plus sûre. L'inaudibilité de cette technique avec ses deux variantes est évaluée en utilisant l'algorithme PEAQ et en calculant le SNR. Nous allons en plus tester la robustesse de chaque variante contre l'attaque: compression / décompression MP3.*

## Mots clefs

Tatouage audio numérique, Modèle psychoacoustique 1, Modèle psychoacoustique 2, MDCT, LSB, Codage Hamming, Inaudibilité, Robustesse, PEAQ, SNR

## 1 Introduction

Le développement de l'internet et l'orientation du monde vers une ère où le numérique prend une place de plus en plus importante facilitent les traitements illégaux sur ces données tels que la copie, le piratage, sans parler des droits d'auteurs qui deviennent de plus en plus non protégés. D'où vient l'intérêt du tatouage numérique [3], dans les années 90, comme une solution efficace contre ce type de menaces. Les techniques de tatouage ne cessent à se multiplier pour trouver un bon compromis entre la

robustesse de la technique vis-à-vis des attaques, l'inaudibilité (il faut que la marque insérée n'altère pas la qualité d'écoute) et la capacité d'insertion.

Dans ce papier nous nous intéressons à la présentation d'une technique de tatouage pour l'audio numérique qui comporte deux variantes selon l'utilisation de l'un des deux modèles psychoacoustiques 1 ou 2 implantés dans les couches 1 ou 3 de la norme de codage audio MPEG.

Ce papier est structuré comme suit : dans la section 2, nous allons détailler les processus d'insertion et de détection de la marque pour les deux variantes de la technique proposée. La section 3 présentera les résultats expérimentaux obtenus. Dans la dernière section, nous clôturons par une conclusion et des perspectives.

## 2 Présentation de la technique de tatouage proposée

D'après une étude détaillée sur le tatouage [4, 5, 6, 7, 8, 9, 10, 11, 12, 13] nous avons pu tirer que le domaine fréquentiel est un bon espace de point de vue robuste et inaudibilité d'où l'idée d'utiliser la MDCT pour passer du domaine temporel au domaine fréquentiel. En plus, dans nos jours, la MDCT est la transformation la plus utilisée dans le codage audio vue qu'elle permet une résolution fréquentielle plus fine tout en étudiant les effets de bord. 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 psychoacoustique de la norme MPEG1 dans la recherche des positions d'insertions. Nous avons essayé en première lieu les caractéristiques du modèle 1 (MPH1 / couche 1) de cette norme pour chercher les positions d'insertion et de ce fait nous avons réalisé une première variante. Par la suite, nous avons pensé à une deuxième variante de la technique en utilisant le modèle 2 (MPH2 / couche 3) de la norme MPEG pour voir l'apport de ce dernier sur les résultats expérimentaux.

Dans la suite, nous allons présenter les deux schémas détaillés d'insertion et de détection pour chaque variante.

## 2.1 Schéma d'insertion de la technique de tatouage en utilisant le MPH1

Le schéma général de l'insertion de la marque est donné par la figure suivante :

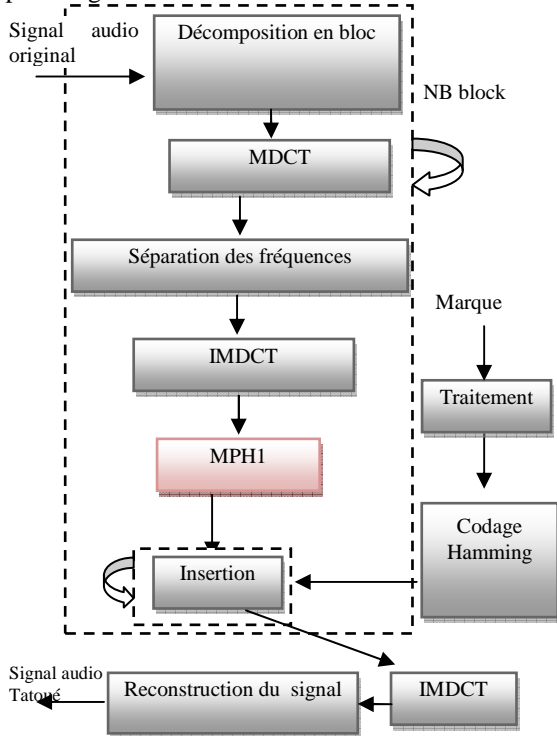


Figure 1 – schéma d'insertion de la marque

La première étape du processus d'insertion de la marque consiste à décomposer le signal audio (. Wav) original en des blocs de taille 1024 échantillons chacun. Par la suite, nous allons appliquer sur chaque sous-bloc temporel (taille 1024) la MDCT pour passer au domaine fréquentiel. Cette transformée va décomposer chaque trame d'échantillons en une bande de basses fréquences (BF) et une bande de hautes fréquences (HF). La séparation de ces deux bandes fréquentielles va être réalisé par un module de séparation des fréquences. La sortie de ce module est représentée sous forme de deux matrices. La première contient les échantillons fréquentiels qui vont constituer l'entrée de l'IMDCT pour être transformés en échantillons temporels : puisque le modèle psychoacoustique de la norme MPEG, (MPH1), s'applique sur des échantillons temporels. Ce dernier (MPH1), va assurer la recherche des positions d'insertion des bits de la marque. Le principe de cette recherche se présente comme suit : après la décomposition du spectre en composantes tonales (représentant les périodicités qui peuvent exister dans le signal) et en composantes non tonales (représentant le bruit qui peut exister dans le signal) et en se basant sur une modélisation du comportement spectral de l'oreille humaine, le modèle va calculer une courbe de masquage au dessous de laquelle

toute insertion sera inaudible. Ensuite, et à partir de ces positions d'insertion recherchées par le MPH1, les bits de tatouage vont être injectés d'une façon substitutive dans la deuxième matrice, représentée par la deuxième sortie du module de séparation des fréquences, qui contient les échantillons fréquentiels non tatoués. Parallèlement, les bits de tatouage vont subir un traitement (binarisation de la marque, décomposition en des portions de 8 bits chacune) puis ils vont être codés selon le principe du codage Hamming (12,8) [14] : puisque les bits de tatouage peuvent subir des changements lors de la détection et de l'insertion de la marque. Le code Hamming permet de détecter les erreurs si elles existent mais de corriger une seule erreur. Chaque bit de la marque, après le traitement nécessaire, va être dupliqué N fois, où N se calcule en fonction du nombre de composantes qui se situent sous la courbe de masquage et de la taille de la marque. Par la suite, chaque bit est inséré dans le bit le moins significatif (LSB) des composantes cherchées par le MPH1.

Toutes les étapes décrites ci-dessus vont être répétées NB bloc fois (avec NB représente le nombre de blocs dans le signal audio). Puis, nous appliquons l'IMDCT sur les blocs fréquentiels tatoués de 1024 échantillons pour obtenir les blocs tatoués dans le domaine temporel. Enfin, la dernière étape consistera à reconstruire le signal audio tatoué résultant.

## 2.2 Schéma de détection de la technique de tatouage en utilisant le MPH1

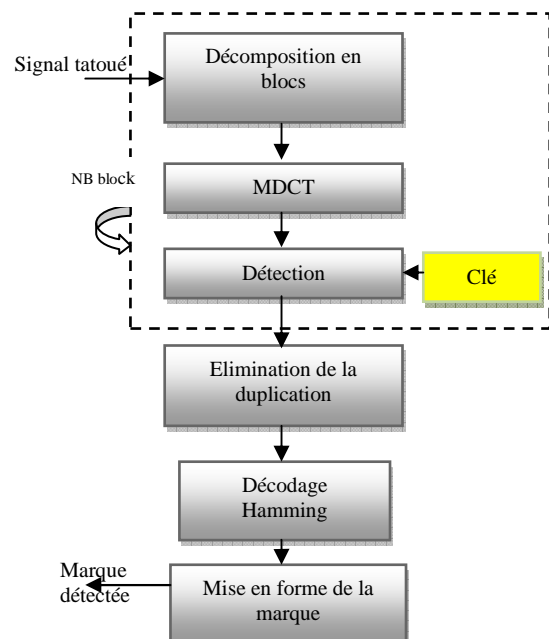


Figure 2 – schéma de détection de la marque

Le processus de détection de la marque est l'inverse de celui de l'insertion. Il ne nécessite pas la présence du signal audio original ni de la marque initialement insérée dans le processus d'insertion. Nous avons besoin seulement de la clé secrète, qui est l'ensemble des positions des composantes qui se situent sous la courbe de masquage cherchée par le modèle psychoacoustique MPH dans la phase de l'insertion, et le nombre de duplication N. Figure 2 illustre les différentes étapes nécessaires pour la détection de la marque.

La sortie du module de détection est un vecteur binaire multiple de 12 (puisque nous avons utilisé un codage de Hamming (12, 8)). L'étape suivante consiste à éliminer la duplication des bits de la marque. Nous procédons par la suite au décodage de Hamming pour détecter et corriger les erreurs si nécessaire. Enfin, nous obtenons une marque décodée (et corrigée) multiple de 8 qui va être mise en forme pour construire la signature détectée finale.

### 2.3 Schéma d'insertion de la technique de tatouage en utilisant le MPH2

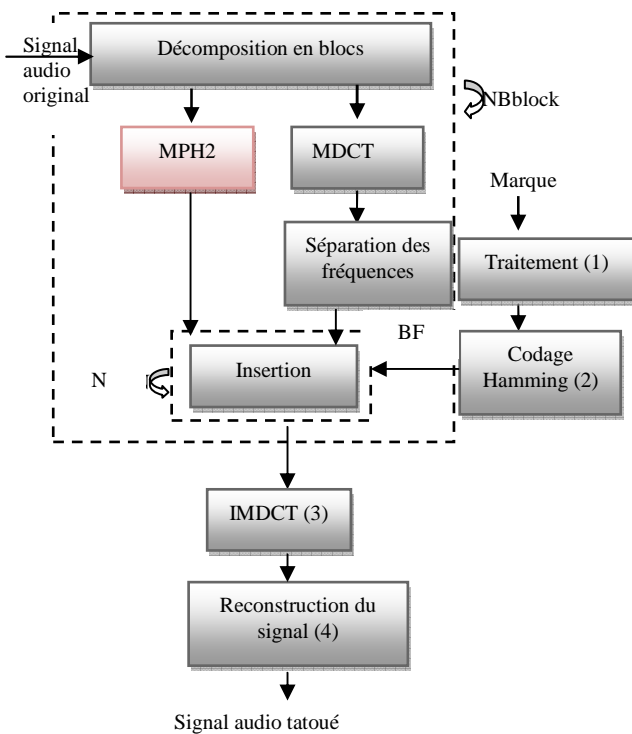


Figure 3 – schéma d'insertion de la marque

La Figure 3 détaille le schéma général et les différentes étapes nécessaires à l'insertion de la marque en utilisant le MPH2 du codeur MPEG1 couche 3.

Dans la première étape, le signal audio original (.Wav) sera divisé en blocs de 1024 échantillons. Par la suite, nous allons appliquer le MDCT pour passer au domaine fréquentiel. Cette transformation va décomposer

chaque bloc en basse fréquence (BF) et haute fréquence (HF). Pour séparer ces deux bandes de fréquences, nous allons utiliser un module de séparation de fréquence. A la fin de cette étape, nous obtenons toutes les basses fréquences où nous allons insérer les bits de la marque. Le choix de la bande des basses fréquences est dû au fait qu'elle est beaucoup moins sensible aux attaques que la bande des hautes fréquences (en particulier contre la compression MP3).

En parallèle et pour rechercher les positions d'insertion les moins sensibles à l'oreille humaine, nous allons appliquer le modèle psychoacoustique 2 (MPH2) de la norme MPEG sur les échantillons temporels de chaque sous-bloc de 1024 échantillons. Les places d'insertion se trouvent sous le seuil de finale de l'audition de l'énergie générée par ce modèle pour chaque bloc. Cette approche offre un bon compromis entre robustesse et inaudibilité. Les étapes de (1) à (4) sont identiques à celles de la première version de la technique.

### 2.4 Schéma de détection de la technique de tatouage en utilisant le MPH2

Toutes les étapes formant le schéma de détection pour cette variante sont les mêmes que celles pour la détection dans la première version (Figure 2), sauf que la clé va se changer. Dans cette version c'est l'ensemble des positions des composantes qui se trouvent sous le seuil de finale de l'audition de l'énergie générée par le modèle 2 pour chaque bloc. A la fin de ce processus, nous obtenons la marque insérée dans le signal original.

## 3 Résultat des tests expérimentaux

Dans cette section, nous allons présenter les différents résultats expérimentaux obtenus par cette technique pour détecter les points forts et les points faibles pour chaque variante.

Ces résultats vont porter sur un corpus expérimental formé de 11 enregistrements. Ces signaux sont échantillonnés en qualité CD (à une fréquence d'échantillonnage  $F_e=44.1$  kHz), de durée 20s en moyenne et sont de style divers : orchestres symphoniques, voix parlées (féminines et masculines), jazz, rock, voix chantée, music arabe. Nous allons insérer la marque texte codée à 136 bits « audiowatermarking », et après codage Hamming de 204 bits (par la suite chaque bit va être dupliqué N fois).

### 3.1 Evaluation objective de la qualité sonore par PEAQ

Pour l'évaluation objective, nous avons utilisé l'algorithme PEAQ [15] qui donne une note d'ODG (Différence grade

Objectif). Cet algorithme compare le signal original et le signal tatoué et attribue une note comprise entre 0 et -4. Le tableau suivant indique la signification de chaque note.

Tableau 1 – Signification des notes d’ODG

Signification	ODG
Imperceptible	0,0
Perceptible mais non gênante	-0,1 à -1
Légèrement gênante	-1,1 à -2
Gênante	-2,1 à -3
Très gênante	-3,1 à -4

Les résultats pour les deux variantes de la technique seront affichés dans la Figure 4. L'axe vertical représente la note donnée par PEAQ.

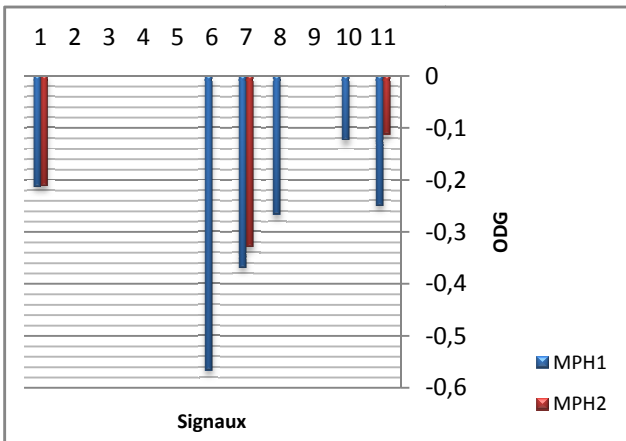


Figure 4 – Représentation graphique des valeurs des notes d’ODG

D’après les résultats affichés dans la figure ci-dessus, nous remarquons que la variante utilisant le MPH2 donne des résultats plus intéressants point de vue inaudibilité.

Pour les extraits 2, 3, 4, 5 et 9 les deux versions de la technique donnent la valeur 0 ce qui signifie que la marque est imperceptible.

Pour les signaux 6, 8 et 10 nous obtenons respectivement les valeurs -0,568, -0,268 et -0,125 avec le MPH1 (perceptible mais non gênante) et 0 (imperceptible) avec le MPH2.

Finalement, pour les extraits 1, 7 et 11 nous obtenons respectivement les valeurs -0,214, -0,37 et -0,251 avec le MPH1 et -0,213, -0,33 et -0,114 avec le MPH2.

### 3.2 Evaluation de la qualité sonore par le calcul du SNR

Une autre façon de montrer l’inaudible de la marque est de calculer le SNR (rapport signal-bruit). Il s'agit d'une mesure qui calcule la similarité entre le son original et le signal audio tatoué.

Les résultats pour les deux variantes de la technique proposée sont présentés dans la figure suivante.

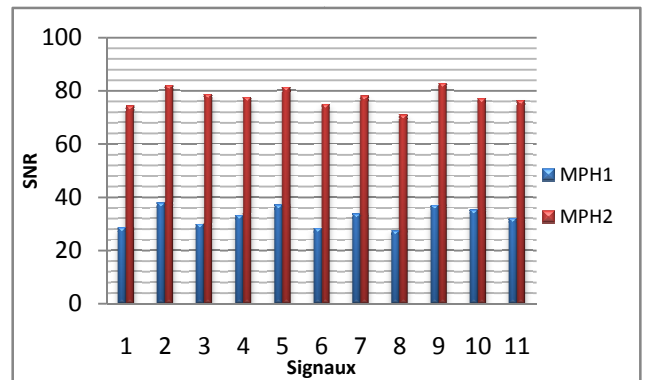


Figure 5 – Représentation graphique des valeurs du SNR

D’après les résultats présentés dans la figure ci-dessus, nous pouvons remarquer par exemple que pour l’enregistrement 2, nous avons atteint une valeur de 81,8938 db avec la variante qui utilise le MPH2 et une valeur de 37,7138 db pour celle qui utilise le MPH1, ce qui prouve encore l’intérêt du MPH2.

### 3.3 Robustesse contre l’attaque compression / décompression MP3

La compression / décompression MP3 est réalisée par "lame.exe" à trois taux différents: 128 kbit / s, 96 Kbit / s et 64 kbit / s. Les résultats des tests sont affichés dans la figure suivante:

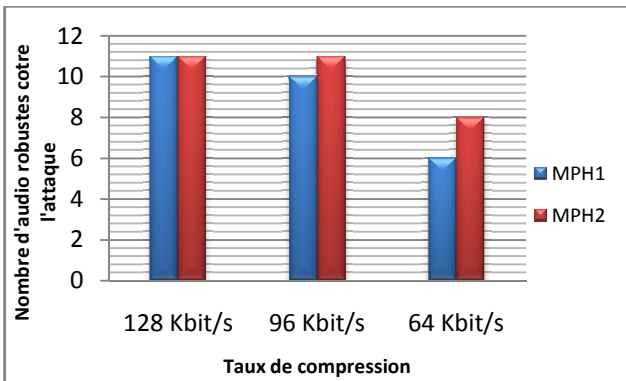


Figure 6 – Robustesse contre l'attaque compression / décompression MP3

Selon les résultats, nous constatons que pour tous les enregistrements, les deux variantes de la technique proposée sont toujours robustes contre l'attaque de compression / décompression pour un taux de 128 Kbit / s. Nous avons pu détecter la marque sans erreur et qui était identique à la marque originale pour tous les signaux. Cette robustesse diminue un peu pour un taux de 96 : nous trouvons 10 signaux robustes contre l'attaque avec le MPH1 alors qu'avec le MPH2 nous trouvons que tous les signaux résistent à cette attaque. A un taux de compression 64 Kbit / s, nous obtenons 6 signaux robustes en utilisant le MPH1 et 8 signaux robustes en utilisant le MPH2.

## 4 Conclusion

Dans cet article, nous avons présenté une technique de tatouage audio numérique dans le domaine fréquentiel avec deux variantes. La première version, permet de chercher les positions d'insertion en utilisant le modèle psychoacoustique 1 de la norme MPEG 1. La deuxième version de la technique utilise les propriétés du modèle psychoacoustique 2 de la norme MPEG1 dans la recherche des places d'insertion. Nous avons présenté par la suite les différents résultats expérimentaux obtenus par chaque variante. Ces résultats montrent l'intérêt du modèle 2 dans l'inaudibilité de la marque malgré la grande capacité d'insertion, ainsi que pour la robustesse contre l'attaque de compression /décompression MP3. Comme perspective, et pour confirmer la robustesse de la technique, avec ses deux variantes, contre l'attaque de compression / décompression MP3 nous visons à calculer le taux d'erreur binaire de la détection pour chaque variante et pour chaque taux de compression.

## Références

- [1] H. Mu-Huo et C. Yu-Hsin. Fast IMDCT and MDCT algorithms— A matrix approach. pg. 221-229. IEEE Trans Signal Process, 2003.
- [2] Norme internationale, ISO/CEI 11172-3. Technologies de l'information codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbit/s, partie 3 : Audio.
- [3] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, Techniques for data hiding. *IBM Systems Journal*, volume. 35, pages. 313–336, 1996.
- [4] Cléo BARAS. Tatouage informé de signaux audio numériques. Thèse de doctorat, Ecole Nationale supérieure des Télécommunications, 2005.
- [5] P. Bassia et I. Pitas. Robust audio watermarking in the time domain. *Proceedings EUSIPCO 98*, volume 1, pages 25-28, Rodos, Greece, 1998.
- [6] L. Boney, A.H Tewfik et K.N. Hamdy. Digital Watermarks for Audio Signals. *IEEE Int. Conf. on Multimedia Computing and Systems June 17-23*, pages 473-480, Hiroshima, Japan, 1996.
- [7] A. Cohen et J. Kovacheckevic. Wavelets: The Mathematical Background. *Proceedings of the IEEE*, Volume 84, No. 4, pages 514-522, 1996.
- [8] Q. Zhang et A. Benveniste. Wavelet Networks. *IEEE Trans. on Neural Networks* Volume 3, No. 6, pages 889-898, 1992.
- [9] N. Cvejic et T.Seppanen. Robust Audio Watermarking in Wavelet Domain Using Frequency Hopping and Patchwork Method. *Proceedings of the 3rd International Symposium on Image and Signal Processing and analysis*, 2003.
- [10] M. Charfeddine, M. El'Arbi, C. Ben Amar. A blind audio watermarking scheme based on Neural Network and Psychoacoustic Model with Error correcting code in Wavelet Domain. *ISCCSP 2008*, pages 12-14, Malta, 2008.
- [11] J. Pinel, L. Girin et C. Baras. Une technique de tatouage haute capacité pour signaux musicaux au format CD-audio. *10 ème congrès français d'acoustique*, pages 12-16, Lyon, Avril 2010.
- [12] D. labarre. Exploitation de la cyclostationnarité pour le tatouage des signaux audio (watermarking), Diplôme d'Etudes Approfondies (DEA). Ecole doctorale des sciences physiques et de l'ingénieur de l'universitéBordeaux I, 2001.
- [13] P. Bassia, I. Pitas. Robust audio watermarking in the time domain. *Congrès signal processing IX: theories and applications*, pages 25-28, 1998.

- [14] Hamming, R. W. Error detecting and error correcting codes. *Bell Syst. Tech. J.*, 26(2), pg. 147–160, 1950.
- [15] Union Internationale des Télécommunications (UIT): Recommandation B.S. 1387: Méthode de mesure objective de la qualité du son perçu, 2001.