

10ème Congrès Français d'Acoustique

Lyon, 12-16 Avril 2010

Une technique de tatouage « haute-capacité » pour signaux musicaux au format CD-audio

Jonathan Pinel¹, Laurent Girin¹, Cléo Baras¹

¹ GIPSA-lab, {jonathan.pinel, laurent.girin, cleo.baras}@gipsa-lab.grenoble-inp.fr

Le tatouage d'un signal (*watermarking* en anglais) consiste à insérer/cacher une information binaire dans un signal de façon imperceptible. Dans le contexte audio qui nous intéresse, ce procédé doit être inaudible. Le tatouage a été initialement proposé et conçu au début des années 1990 pour des applications de protection des droits des œuvres numériques (DRM pour *Digital Rights Management* en anglais). Dans ce cas, l'effort s'est porté sur la robustesse du procédé vis-à-vis d'attaques visant à le neutraliser, au dépend du débit d'informations insérées (de l'ordre de la dizaine de bits/s pour l'audio). Depuis, le tatouage audio est utilisé pour d'autres applications, comme le transport de métadonnées liées au signal porteur. Cependant, les débits restent généralement assez faibles, et dans cette étude nous proposons une technique de tatouage audionumérique « haute-capacité » pour des signaux au format CD-audio. La technique proposée repose sur une quantification QIM (*Quantization Index Modulation*) des coefficients MDCT (*Modified Discrete Cosine Transform*) du signal avec prise en compte de contrainte perceptive par un modèle psycho-acoustique assurant l'inaudibilité du tatouage. Nous montrons que cette technique peut atteindre des débits de l'ordre de 250kbits/s par canal (selon le type de signal audio porteur), au détriment de la robustesse : elle peut être utilisée dans des applications « non-sécuritaires » où l'on suppose que le signal ne subit pas d'attaques intentionnelles (autres que la conversion au format CD-audio dont l'effet est explicitement présenté dans l'étude). En l'occurrence, nous utilisons cette technique dans un système de séparation de sources informée présenté dans ce même congrès.

1 Introduction

Le tatouage audionumérique, qui consiste à insérer de l'information de façon indétectable (ici inaudible) dans un media, est apparu au début des années 1990 [1][2][3]. Il a pendant longtemps été utilisé à des fins purement sécuritaires, dans le cadre de la Gestion des Droits Numériques. En effet, grâce aux améliorations des supports de stockage et des techniques de compression, et surtout au développement d'internet, le « piratage » s'est accru. Le tatouage sécuritaire consiste à insérer une information de faible taille (comme une signature numérique) dans le signal audio. La difficulté pour cette utilisation n'est donc pas le débit de tatouage, qui peut être assez faible, mais la robustesse du tatouage à de possibles attaques ayant pour but de neutraliser le message inséré.

Depuis plusieurs années, le tatouage a tendance à s'étendre à d'autres domaines que celui de la sécurité : on utilise le tatouage non plus comme un moyen de transporter une information totalement cachée aux yeux de l'utilisateur, mais comme un porteur d'information additionnelle qui pourra être récupérée par l'utilisateur [4] (on parle alors de contenu augmenté). Par exemple, dans [5][6], on utilise le tatouage pour transmettre une information permettant de séparer les sources composant un mélange tatoué. Pour de nombreuses applications potentielles, le point important n'est donc plus tant la robustesse aux attaques (puisque l'utilisateur n'a aucun intérêt à dégrader l'information) que le débit : on souhaite insérer un maximum d'information dans le signal

audio de façon inaudible.

Dans ce papier, on considère un tel cadre applicatif et nous présentons une technique de tatouage haute-capacité développée pour les signaux audio au format CD (44,1 kHz, stéréo, codage PCM 16 bits des échantillons temporels), dont le but est de maximiser le débit de tatouage sous contrainte d'inaudibilité.

Ce papier est organisé de la façon suivante : la section 2 est une présentation générale du système de tatouage et la section 3 une présentation plus détaillée des blocs principaux du système. Des résultats sont présentés en section 4 et la section 5 conclut cet article.

2 Vue d'ensemble du système de tatouage

Dans cette section nous présentons rapidement le système de tatouage dans son ensemble ; chaque partie sera détaillée ensuite dans la section 3. Le système (voir figure 1 pour le schéma) est constitué de deux blocs principaux : un codeur (figure 1a) et un décodeur (figure 1b). Au niveau du codeur, la technique de tatouage utilisée s'effectuant dans le domaine temps-fréquence, le signal d'entrée x (au format CD) est tout d'abord transformé dans le plan temps-fréquence par la MDCT¹ (*Modified Discrete Cosine Transform*) au bloc ①. Les coeffi-

¹La MDCT implique un découpage du signal en trames indicées par t et en canaux fréquentiels indicées par f . Les traitements sont effectués à l'échelle d'un bloc, et on omet l'indice t pour les traitements internes à un bloc.

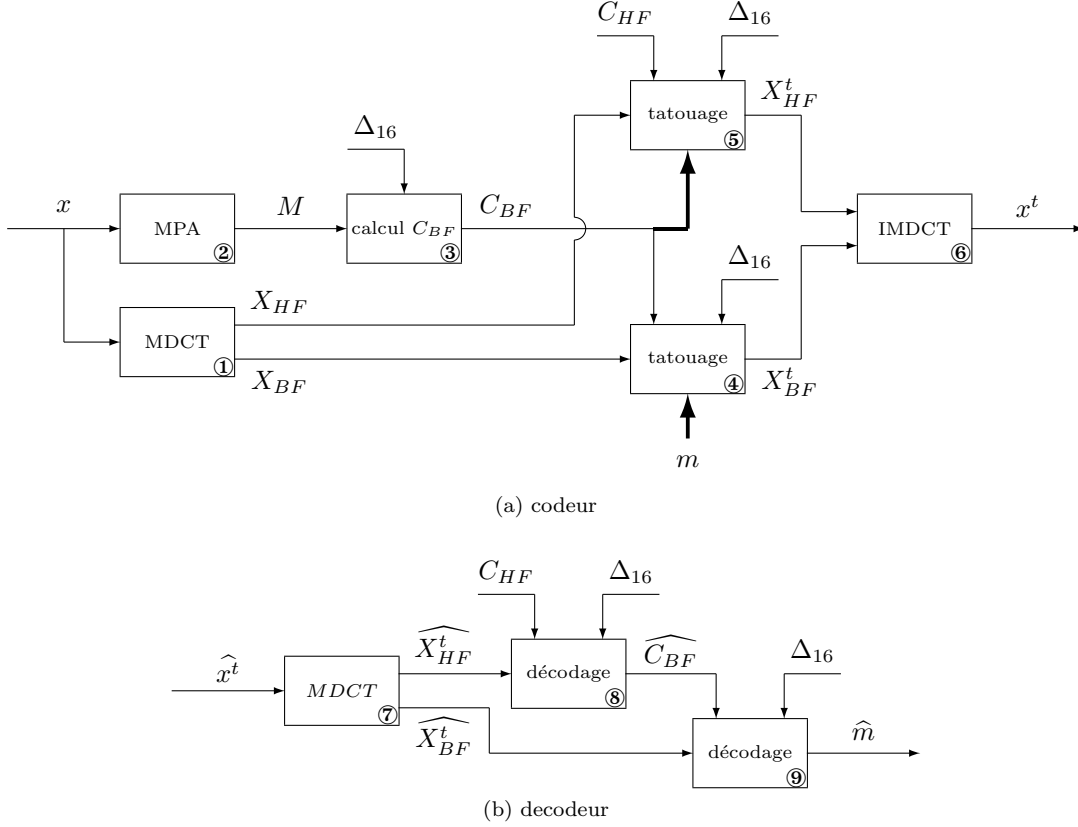


FIG. 1: Schémas du codeur et décodeur. La notation $\hat{\cdot}$ indique la possible dégradation du signal par la quantification CD. Les flèches en trait épais indiquent une information insérée par tatouage.

cients MDCT sont ensuite séparés en une partie « basses fréquences » ($_{BF}$ sur le schéma) occupant la majeure partie du spectre, et une partie « hautes fréquences » ($_{HF}$ sur le schéma) limitée aux canaux les plus élevés. Cette séparation constitue un point central de notre système de tatouage : le modèle psychoacoustique (bloc ②) et son seuil de masquage $M(f)$, inspirés du modèle de la norme MPEG-AAC permettent de calculer, pour chaque canal fréquentiel f de la zone basses fréquences, la capacité de tatouage (*i.e.* la taille du maximale code binaire à insérer) $C_{BF}(f)$ sous contrainte d'inaudibilité (bloc ③). Les hautes fréquences sont quant à elles utilisées pour tatouer les valeurs des capacités $C_{BF}(f)$ qui paramétrisent le tatouage des basses fréquences. Il y a donc deux opérations de tatouage au codeur, la première étant le tatouage du message m dans les basses fréquences aux capacités $C_{BF}(f)$ (bloc ④) et la deuxième le tatouage des valeurs de ces capacités $C_{BF}(f)$ dans les hautes fréquences (bloc ⑤). Pour cela, on impose des capacités fixes $C_{HF}(f)$ sur les coefficients hautes fréquences. La technique de tatouage utilisée dans les deux cas est la QIM (*Quantization Index Modulation*), bien que l'on puisse envisager des techniques différentes pour les deux zones de tatouage. Une fois les coefficients MDCT tatoués, le signal est tout simplement repassé dans le domaine temporel par la transformée inverse de la MDCT, l'IMDCT (*Inverse Modified Discrete Cosine Transform*) au bloc ⑥.

Au niveau du décodeur, le signal temporel tatoué reçu $\hat{x}^t$ est tout d'abord passé dans le domaine temps-fréquence (bloc ⑦) et séparé en basses et hautes fréquences comme au codeur. Les capacités $C_{HF}(f)$ des

hautes fréquences étant fixes, on décode tout d'abord l'information tatouée dans cette zone (bloc ⑧) : ce sont les valeurs décodées $\widehat{C_{BF}}(f)$. Cette information est ensuite utilisée pour décoder le message $\hat{m}$ au bloc ⑨ et donc récupérer l'information « utile ».

Une particularité de ce système de tatouage est que l'on peut choisir la taille N de la fenêtre pour traiter un signal. Cette variabilité est guidée par deux points : d'abord cette taille de fenêtre N est un paramètre susceptible de faire varier les performances du tatouage du système, et sera testée en tant que telle. Ensuite, ce système peut être utilisé conjointement avec des applications qui utilisent la décomposition MDCT ; il peut donc être intéressant d'avoir la même taille N pour l'application et le système de tatouage.

3 Présentation détaillée

Dans cette section nous décrivons en détails les blocs ou techniques composant le système de tatouage.

3.1 Transformée temps-fréquence

Du fait de la nécessité d'un haut débit, la technique de tatouage est basée sur des techniques de quantification. Or quantifier directement les échantillons temporels conduit vite à une dégradation de la qualité du signal. C'est pourquoi, comme déjà mentionné dans la section 2, le tatouage est effectué sur des coefficients temps-fréquence du signal. En l'occurrence, le choix de la MDCT (*Modified Discrete Cosine Transform*, blocs

① et ⑦, et bloc ⑥ pour la MDCT inverse) à été guidé par plusieurs points :

- les coefficients MDCT sont réels (à l’opposé des coefficients complexes pour la FFT)
- ils sont particulièrement robustes à la quantification (et plus généralement à des manipulations modifiant leurs valeurs)
- la MDCT est une transformée à recouvrement, qui donne d’excellents résultats contre le repliement temporel (on parle de TDAC : *Time-Domain Aliasing Cancellation*).

C’est pour ces raisons notamment que la MDCT est utilisée dans de nombreuses applications pour l’audio, en particulier en compression (AAC, AC3, Vorbis...).

D’un point de vue technique, l’expression de la MDCT appliquée sur des blocs de N échantillons du signal temporel est la suivante :

$$X_M(f) = \sum_{n=0}^{N-1} x(n)w_a(n) \cos\left(\frac{2\pi}{N}(n+n_0)\left(f+\frac{1}{2}\right)\right),$$

avec $f \in [0, \frac{N}{2}-1]$, $n_0 = \frac{1}{2} + \frac{N}{4}$ et w_a la fenêtre d’analyse. La formule de la transformée inverse (IMDCT) est la suivante :

$$y(n) = w_s(n) \frac{4}{N} \sum_{k=0}^{\frac{N}{2}-1} X_M(f) \cos\left(\frac{2\pi}{N}(n+n_0)\left(f+\frac{1}{2}\right)\right),$$

avec $n \in [0, N-1]$ et w_s la fenêtre de synthèse.

On remarque ici que la MDCT de N échantillons du signal ne donne que $\frac{N}{2}$ coefficients. Ce n’est donc pas à proprement parler une transformée inversible. Cependant en prenant un recouvrement des blocs successifs de 50% et si les fenêtres d’analyse w_a et de synthèse w_s vérifient certaines conditions, on a une reconstruction parfaite du signal original. On peut trouver plus de détails sur cette transformée dans [7].

3.2 Modèle Psychoacoustique

Le modèle psychoacoustique utilisé dans notre système (bloc ②) est directement inspiré du modèle psychoacoustique de MPEG-AAC [8][9], avec quelques adaptations pour permettre une taille de fenêtre d’analyse N variable. Dans AAC, le principe général est de calculer un rapport signal à masque (SMR pour *Signal to Mask Ratio*) qui guide l’allocation des bits de codage dans les différentes sous-bandes (par l’intermédiaire d’un calcul d’entropie perceptuelle; cette dernière déterminant par ailleurs la taille de la fenêtre d’analyse, 2048 ou 256). Dans notre cas, nous allons nous servir de ce rapport signal à masque afin d’obtenir une courbe de masquage globale pour chaque bloc de signal permettant de mettre en forme le bruit de *tatouage* (sous cette courbe) de manière à ce qu’il soit inaudible. C’est ce seuil de masquage qui permet donc de déterminer les capacités de tatouage. On peut noter que si l’on a besoin d’un débit de tatouage particulièrement élevé ou faible, il est possible d’ajuster ce masque avant le calcul des capacités (en le multipliant par une constante en linéaire ou en effectuant un décalage en dB). On peut donc gagner en débit au prix d’une perte de qualité ou inversement.

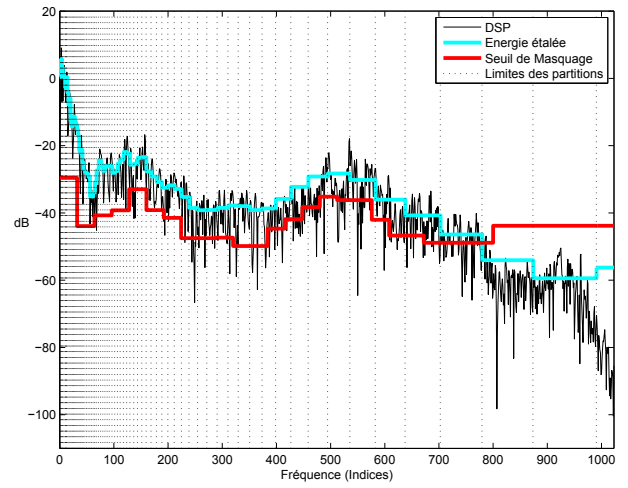


FIG. 2: Exemple de grandeurs du modèle psychoacoustique pour une fenêtre de taille $N=2048$.

Le modèle psychoacoustique AAC fonctionne à partir du spectre FFT du signal, que l’on étale par convolution avec une fonction d’étalement. On ajuste alors cette énergie étalée en dB suivant la tonalité du signal et finalement on procède à un contrôle du pré-écho et à la prise en compte du seuil d’audition absolu, ce qui donne le seuil de masquage final (voir figure 2 pour un exemple).

Une caractéristique importante du modèle psychoacoustique AAC est que toutes les grandeurs intermédiaires sont définies non pas par coefficient fréquentiel f mais par « partition », le rapport signal à masque étant quant à lui défini par bande de codage. Dans AAC, les partitions sont approximativement égales au minimum entre un tiers de bande critique (en échelle Bark [10]) et un canal fréquentiel, afin de garder un modèle psychoacoustique de bonne qualité. Les bandes de codage sont quant à elles fixées arbitrairement (leur paramétrage ne modifie pas les principaux calculs du modèle psychoacoustique). La norme AAC n’utilisant que deux tailles de fenêtre (2048 et 256), ces partitions sont inscrites dans des tables. Afin d’assurer l’adaptabilité de notre système au niveau de la taille N des fenêtres, nous avons donc développé un algorithme de calcul des partitions en fonction de N , N pouvant prendre les valeurs des puissances de 2 (dans notre cas nous nous intéressons plus particulièrement aux valeurs 512, 1024, 2048 et 4096). Cet algorithme calcule simplement les partitions en partant de l’indice fréquentiel 0 afin que la largeur des partitions soit la plus proche possible d’un tiers de bande critique, en utilisant l’expression analytique pour la conversion des Barks en Hertz donnée dans [11] avec la contrainte qu’une partition soit au moins égale à 1 canal fréquentiel.

Pour ce qui est des bandes de tatouage (qui correspondent aux bandes de codage AAC), leur nombre et leur taille doivent être déterminés. Leur nombre ne peut pas être trop élevé : en effet, pour les basses fréquences, la capacité étant fixée dans chaque bande de tatouage, il faudra donc tatouer dans les hautes fréquences autant de valeurs de capacités qu’il y a de bandes de tatouage dans les basses fréquences. Nous avons choisi deux types de répartition pour les bandes de tatouage :

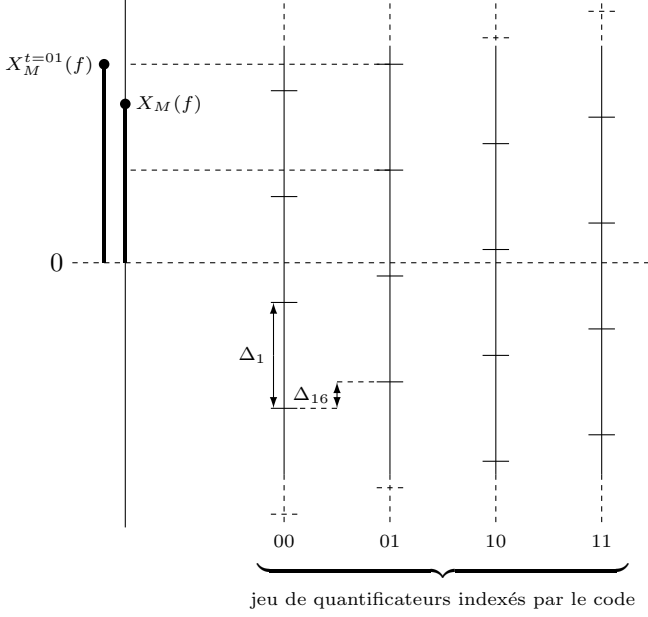


FIG. 3: Positionnement des grilles de quantification QIM. On veut tatouer le code 01 au coefficient MDCT $X_M(f)$, on le quantifie donc en $X_M^{t=01}(f)$ sur le quantificateur indexé par 01.

- une avec 25 bandes réparties suivant les bandes critiques (Barks, ayant donc une répartition à peu près logarithmique),
- une avec 32 bandes réparties uniformément (comme dans le banc de filtre de MPEG-1 [12]).

3.3 Tatouage

3.3.1 QIM

La technique de tatouage utilisée est la QIM (*Quantization Index Modulation* [13]), une technique de tatouage par quantification (tatouage aux blocs ④ et ⑤, décodage aux blocs ⑧ et ⑨). Le principe consiste à se munir (pour chaque coefficient MDCT) de jeux de quantificateurs entrelacés, indexés par les codes possibles à tatouer. Pour une capacité $C(f)$, on obtient un jeu de $2^{C(f)}$ quantificateurs pour le coefficient MDCT $X_M(f)$. Ces grilles de quantification sont ici réparties uniformément comme on peut le voir sur l'exemple figure 3, afin d'optimiser les performances d'inaudibilité et de décodage du tatouage. L'opération de tatouage proprement dite est faite en quantifiant le coefficient MDCT sur la grille indexée par le code que l'on veut tatouer (voir exemple figure 3 pour une capacité de 2 bits). Au codeur, un coefficient est remplacé par sa valeur quantifiée la plus proche sur la grille correspondant au message à transmettre. Au décodeur, le message est récupéré en estimant de quelle grille de quantification le coefficient MDCT $X_M^t(f)$ obtenu à partir du signal tatoué est le plus proche.

Il reste alors à déterminer les capacités $C(f)$, qui seront calculées à partir du seuil de masquage $M(f)$

pour les basses fréquences ($C_{BF}(f)$) et fixées pour les hautes fréquences ($C_{HF}(f)$), comme expliqué dans la sous-section suivante.

3.3.2 Calcul et choix des capacités

Le calcul des capacités en basses fréquences $C_{BF}(f)$ se fait au bloc ③. On veut maximiser le débit de tatouage en suivant le modèle psychoacoustique, cela correspond donc à un problème d'optimisation sous double contrainte (robustesse à la quantification CD et inaudibilité). Dans la plupart des applications utilisant la quantification (telles que la compression), on transmet en plus du signal codé les paramètres des quantificateurs pour le décodage (généralement les résolutions et les facteurs d'échelle). Ici la quantification ne sert pas à la transmission mais au tatouage du signal, il n'y a pas besoin de transmettre tous les paramètres des quantificateurs, car pour les reconstruire au décodeur on a besoin uniquement des capacités par canal fréquentiel $C(f)$.

Robustesse à la quantification CD : Si le système n'a pas pour but d'être robuste à des attaques, il faut tout de même qu'il soit robuste à la quantification CD des échantillons temporels du signal tatoué, ce qui revient à ajouter un bruit sur les coefficients MDCT. Le théorème central limite permet de considérer que l'effet de ce bruit est très proche d'un bruit blanc additif gaussien d'écart-type $\sigma_{16} = \sqrt{N} \cdot 4,5 \cdot 10^{-6}$. Nous pouvons donc choisir une probabilité d'erreur p_e (10^{-6} dans notre système) et calculer l'écart minimum Δ_{16} entre deux valeurs de tatouage (voir figure 3) pour avoir un taux d'erreur de relecture du message inférieur à p_e . Le pas de quantification $\Delta_1(f)$ des quantificateurs pour le coefficient MDCT d'indice f est alors donné par :

$$\Delta_1(f) = 2^{C(f)} \cdot \Delta_{16}. \quad (1)$$

Inaudibilité : La contrainte d'inaudibilité est guidée dans les basses fréquences par le seuil de masquage $M(f)$ en sortie du modèle psychoacoustique. La contrainte est que l'erreur de tatouage dans le pire des cas se trouve sous le seuil de masquage, c'est-à-dire :

$$\left(\frac{\Delta_1(f)}{2} \right)^2 < M(f), \quad (2)$$

pour toute fréquence f des bandes de tatouage des basses fréquences. Dans les hautes fréquences, ayant besoin de capacités fixes, la contrainte est que ces capacités $C_{HF}(f)$ soient assez faibles pour être toujours inaudibles quelle que soit la trame actuelle (cela revient à considérer uniquement le seuil d'audition absolu).

Choix des capacités : En combinant les expressions (1) et (2), on obtient pour les basses fréquences :

$$C_{BF}(f) = \left\lfloor \frac{1}{2} \log_2 \left(\frac{M(f)}{\Delta_{16}^2} \right) + 1 \right\rfloor.$$

Les valeurs obtenues pour ces capacités basses fréquences sont, d'après nos expériences, toujours inférieures à 15. On peut donc coder leurs valeurs avec

Absolute Grade	5,0	Imperceptible	0,0	ODG
	4,9 à 4,0	Perceptible mais pas gênante	-0,1 à -1	
	3,9 à 3,0	Légèrement gênante	-1,1 à -2	
	2,9 à 2,0	Gênante	-2,1 à -3	
	1,9 à 1,0	Très gênante	-3,1 à -4	

TAB. 1: Signification des notes d'ODG.

un code de taille fixe sur 4 bits (de 0 à 15). Afin d'être sûr à la fois de respecter la contrainte d'inaudibilité pour les hautes fréquences et d'avoir une zone hautes fréquences de taille réduite, nous avons fixé les capacités $C_{HF}(f)$ à 1 ou 2 bits par coefficient MDCT². Pour le découpage des bandes de tatouage suivant l'échelle des Barks (25 bandes espacées de manière quasi-logarithmique), la dernière bande étant très large, on prend pour hautes fréquences les 50 derniers canaux fréquentiels que l'on tatoue à 2 bits par coefficients. Cela donne 100 bits, ce qui permet bien de coder les 25 valeurs des capacités des bandes de tatouage sur 4 bits. Pour le découpage en 32 bandes de même taille, on prend pour hautes fréquences les n_{HF} dernières bandes, avec n_{HF} le plus petit entier tel que la capacité de tatouage de ces n_{HF} sous-bande (toujours avec $C_{HF}(f) = 1$ ou 2) permette de coder les valeurs des capacités $C_{BF}(f)$ des $32 - n_{HF}$ premières bandes de tatouage. A titre d'exemple, pour des fenêtres de taille $N = 2048$, pour la répartition en 32 bandes cela correspond à 2 des 32 bandes (donc 64 coefficients contre 50 pour la répartition des bandes suivant l'échelle des Barks).

4 Résultats

Afin de tester notre système de tatouage, nous avons utilisé une dizaine d'extraits de morceaux de musique de styles différents d'une durée de 10 secondes chacun environ. Nous nous sommes intéressés pour chacun de ces extraits à la qualité du son et au débit de tatouage. Comme déjà mentionné dans la section 3.2, des tests ont été effectués pour deux types de bande de tatouage, une avec 32 bandes réparties uniformément (*lin*) et une avec 25 bandes réparties approximativement suivant les bandes critiques (*log*).

4.1 Inaudibilité du tatouage

La qualité sonore du signal tatoué a été évaluée via l'algorithme PEAQ (*Perceptual Evaluation of Audio Quality* [14]), qui donne une note d'ODG (*Objective Difference Grade*). Cet algorithme compare le signal original et le signal tatoué et renvoie une note comparative entre 0 et -4 dont les significations sont présentées tableau 1.

²la marge étant très élevée, il est envisagé dans des travaux futurs de faire des tests avec des capacités plus élevées pour la zone hautes fréquences afin de gagner de la place pour l'information utile.

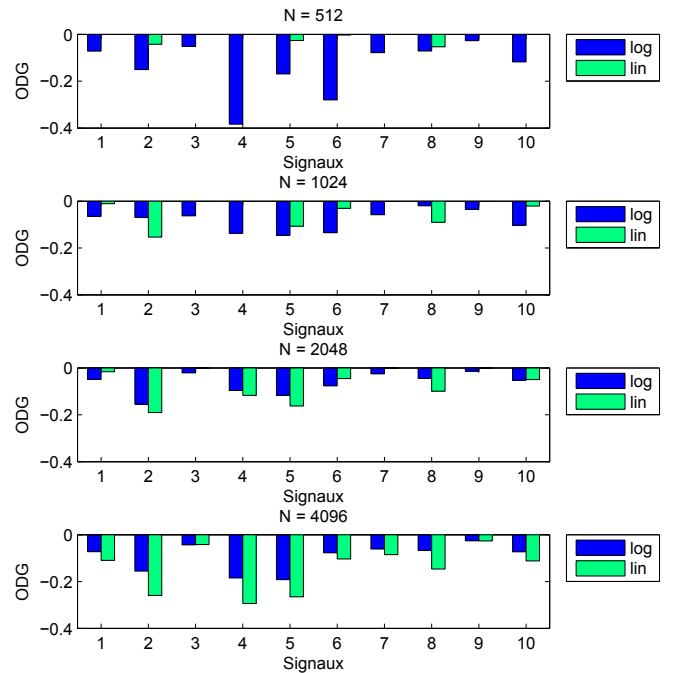


FIG. 4: Notes d'ODG et des morceaux utilisés suivant la tailles des fenêtres N .

Les notes d'ODG obtenues pour les deux variantes de notre système sont présentées figure 4. On remarque que quel que soit le schéma de répartition des bandes de tatouage, les notes d'ODG obtenues sont bonnes, puisqu'aucune n'est en dessous de -0.5, et même très bonnes pour des tailles de fenêtres supérieures à 512 avec des minima entre -0.2 et -0.3. Nous sommes donc assurés que notre système ne dégrade que très peu la qualité des signaux musicaux. Si le système avec la répartition *log* semble de moins bonne qualité pour des tailles de fenêtres de 512, cette tendance paraît s'inverser petit à petit lorsque l'on augmente la taille des fenêtres.

4.2 Débits d'insertion

Les débits moyens obtenus sont présentés pour chaque taille de fenêtre utilisée en fonction des morceaux de musique figure 5a et en moyenne sur les morceaux en fonction de la taille des fenêtres figure 5b. On remarque ici que même si les débits obtenus sont assez variables suivant les styles musicaux, ils n'en restent pas moins élevés, avec une moyenne dépassant les 200kbts/s quelle que soit la taille de la fenêtre utilisée. Une moyenne de 250kbts/s est atteinte pour des tailles de fenêtre $N = 2048$ (taille standard pour beaucoup de systèmes utilisant la MDCT ; c'est notamment la taille des fenêtres « longues » de AAC). Le débit d'un canal audio au format CD (44.1kHz, codage PCM 16bits des échantillons temporels) étant de 705kbts/s, cela représente environ 30 à 35% du débit du signal audio. On observe aussi une nette augmentation du débit lorsque la taille des fenêtres augmente. Ceci est lié à deux points :

- le modèle psychoacoustique étant basé sur le domaine fréquentiel, plus on améliore la résolution dans ce domaine (c'est-à-dire plus la taille des fenêtres est élevée), meilleurs seront les résultats.

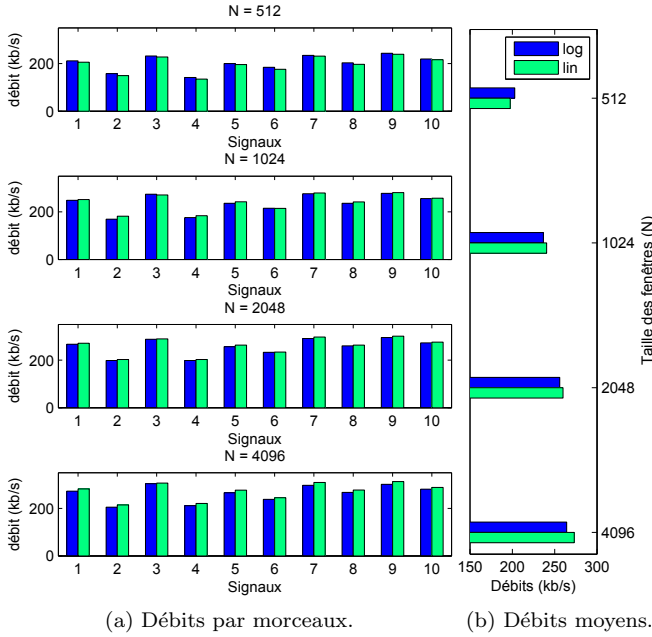


FIG. 5: Débits obtenus pour chacun des morceaux et moyennes suivant la tailles des fenêtres N .

Cependant, il faut bien faire attention à toujours respecter la dynamique du signal et donc à ne pas prendre des fenêtres trop grandes (de l'ordre de 50ms).

- le nombre de bits utilisés dans les hautes fréquences pour tatouer les valeurs des capacités $C_{BF}(f)$ étant quasiment invariant suivant la taille de fenêtre N , lorsque l'on augmente cette taille de fenêtre (et donc que l'on diminue le nombre de trames), on réduit le nombre total de bits de tatouage en hautes fréquences; ce qui permet de gagner du débit pour l'information utile.

5 Conclusions et perspectives

La technique de tatouage présentée dans ce papier permet de tatouer des signaux musicaux à très haut débit (plus de 200kb/s) tout en gardant une très bonne qualité et en étant robuste à la quantification CD.

Les perspectives envisagées sont multiples :

- mieux prendre en compte le phénomène de pré-écho, en effet c'est lui qui est responsable des notes d'ODG relativement plus faibles de certains morceaux pour les tailles de fenêtres faibles (512 essentiellement);
- améliorer la répartition des bandes de tatouage pour gagner en débit ou en qualité;
- adapter le modèle psychoacoustique afin de pouvoir recalculer les capacités au décodeur sans avoir à coder des valeurs de capacités dans les hautes fréquences. Ceci permettrait de libérer de la capacité pour l'information utile.

Remerciements

Ces travaux sont soutenus par l'Agence Nationale de la Recherche française (ANR) dans le cadre du projet

Références

- [1] M. Costa. Writing on dirty paper. *IEEE Trans. Inform. Theory*, 29(3) :439–441, 1983.
- [2] L. Boney, T. Ahmed, and H. Khaled. Digital watermarks for audio signals. *Third IEEE International Conference on Multimedia Computing and Systems*, pages 473–480, June 1996.
- [3] I.J. Cox, M.L. Miller, and A.L. McKellips. Watermarking as communications with side information. *Proc. IEEE*, 87(7) :1127–1141, 1999.
- [4] B. Chen and C.-E.W. Sundberg. Digital audio broadcasting in the FM band by means of contiguous band insertion and precanceling techniques. *IEEE Trans. Commun.*, 48(10) :1634–1637, 2000.
- [5] M. Parvaix, L. Girin, and J.-M. Brossier. A watermarking-based method for single-channel audio source separation. In *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pages 101–104, Taipei, Taiwan, 2009.
- [6] M. Parvaix and L. Girin. Séparation de source informée pour des mélanges stéréo instantanés utilisant un tatouage de l'index des sources localement prédominantes. In *Actes du 10ème Congrès Français d'Acoustique*, Lyon, 2010. Société Française d'Acoustique.
- [7] J.P. Princen and A.B. Bradley. Analysis/synthesis filter bank design based on time domain aliasing cancellation. *IEEE Trans. Acoust., Speech, Signal Process.*, 64(5) :1153–1161, 1986.
- [8] ISO/IEC. *ISO/IEC 13818-7 :2004(E) Information technology - Generic coding of moving pictures and associated audio information - Part 7 : Advanced Audio Coding (AAC)*.
- [9] N. Moreau, O. Derien, S. Larbi, and M. Perreau Guimares. Le codeur MPEG-2 AAC expliqué aux traités de signaux. *Ann. Télécommun.*, 55(9-10) :442–461, 2000.
- [10] E. Zwicker and U. Zwicker. *Psychoacoustics Facts and Models*. Springer-Verlag, 1990.
- [11] H. Traunmüller. Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am.*, 88 :97–100, 1990.
- [12] ISO/IEC MPEG. *IS11172-3 Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbps/s, Part 3 : Audio*. ISO, 1992.
- [13] B. Chen and G. Wornell. Quantization index modulation : A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Inform. Theory*, 47(4) :1423–1443, 2001.
- [14] ITU. *ITU-R Recommendation BS.1387-1 : Method for objective measurements of perceived audio quality (PEAQ)*, 2001.