Modulations par étalement de spectre dans un système de tatouage audio

C. Baras N. Moreau

Département Traitement du Signal et des Images

Ecole Nationale Supérieure des Télécommunications
46 rue Barrault, 75013 Paris, FRANCE

{baras, moreau}@tsi.enst.fr

Résumé

L'une des applications du tatouage consiste à utiliser le signal audio comme un canal de transmission porteur d'information. Débit et fiabilité de transmission deviennent alors des objectifs primordiaux. On se propose d'étudier l'apport de différentes modulations par étalement de spectre sur les performances d'un système de tatouage, présentement décrit. Nous en déduirons la modulation la plus adaptée dans le cas d'une perturbation du canal réalisée par un codeur MPEG.

Mots clefs

Tatouage, audio, débit, TEB, CDMA, Gold.

1 Introduction

Avec le développement croissant des échanges de fichiers audio sous format numérique, la protection des droits de propriété intellectuelle est devenue un problème majeur. Le tatouage s'est imposé comme une solution potentielle à ce problème depuis quelques années. Il consiste à insérer une marque indélébile directement dans le signal audio, satisfaisant généralement aux contraintes suivantes : l'inaudibilité, la robustesse aux opérations de traitement classique sur les signaux et la résistance à des attaques délibérées de tiers. Ces études ont permis d'étendre le tatouage à de nouveaux domaines d'application : l'un d'eux consiste à utiliser le signal audio comme un canal de transmission véhiculant une information binaire. Le système de tatouage se présente alors comme une chaîne de communication aux propriétés très particulières. L'information utile, le tatouage, est transmise avec une faible puissance devant un bruit, le signal audio, fortement corrélé et non stationnaire. Si cette application s'affranchie de la notion de "pirate", elle doit néanmoins respecter les contraintes classiques d'inaudibilité du tatouage (assuré par les propriétés de masquage du système auditif humain) et de robustesse (notamment à des opérations de compression) et surtout satisfaire à des objectifs de débit et de fiabilité de transmission. De fait le couple débit - taux d'erreur binaire (TEB) définit la notion de performance d'un système de tatouage dédié à la transmission d'information. A ce couple s'ajoute également le facteur coût en terme de temps de calcul, dont la prise en compte est nécessaire lors de l'implémentation du système de tatouage.

Un état de l'art du tatouage dans le domaine de l'audio est présenté dans le premier chapitre de la thèse de L. de C. T. Gomes [1].

On se propose d'étudier l'apport de différentes modulations par étalement de spectre sur les performances d'un système de tatouage à vocation de transmission de données [2]. Après une description du système de tatouage utilisé, nous détaillerons les techniques de modulation envisagées. Une analyse des résultats expérimentaux de ces modulations en terme de débit, de TEB et de temps de calcul nous permettra de déterminer la meilleure modulation à choisir pour un tel système.

2 Principe du système de tatouage utilisé

Mis en oeuvre en collaboration avec L. de C. T. Gomes dans le cadre de sa thèse [1], ce système se présente sous la forme d'une chaîne de communication, dont le schéma de principe est donné figure 1. L'information binaire est codée à l'aide d'un dictionnaire $\mathcal{D} = \{\underline{d}, -\underline{d}\}$ associant à chaque bit un signal blanc gaussien (ou vecteur) de durée N $\underline{d} = [d(0)...d(N-1)]$ et de puissance unité. Le choix du signe de $\underline{d}$ est fonction du bit à émettre. Ce signal est ensuite filtré de sorte que sa densité spectrale de puissance s'approche au mieux du seuil de masquage, limite fréquentielle caractérisant la contrainte d'inaudibilité. Ce seuil et implicitement le filtre, sont calculés à l'aide d'un modèle psychoacoustique (similaire à ceux utilisés pour les opérations de compression) et sont actualisés sur chaque fenêtre d'analyse du signal audio à tatouer. L'information binaire ainsi mise en forme est finalement insérée par sommation temporelle avec le signal audio numérique échantillonné à F_e.

La perturbation du canal par une compression de type MPEG entraîne une suppression des hautes fréquences du signal tatoué. La théorie des transmissions sur canal à bande passante limitée [3] préconise l'introduction à l'émetteur d'un filtre passe-bas, de fréquence de coupure

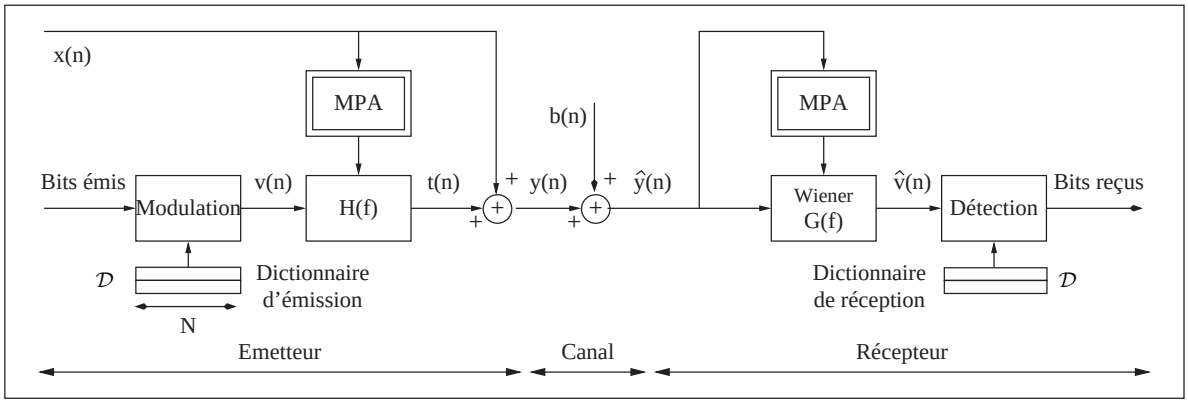

Figure 1 – *Schéma du système de tatouage*

approchant celle du canal : ce filtre prend effet au niveau du dictionnaire, qui est donc constitué de signaux blancs limités dans la bande de fréquence $[0, F_c]$, et du filtre de mise en forme $H(f)$. Travaillant sur des signaux audio échantillonnés à 44.1 kHz, le choix a été fait, après analyse des spectrogrammes des signaux audio compressés, de limiter la bande fréquentielle à $F_c = 11$ kHz.

La réception exploite un filtrage de Wiener, qui réalise à la fois l'égalisation du canal et le blanchiment du signal audio tatoué. Il en résulte une estimation du signal modulé, qui est alors soumis à un détecteur par corrélation. La corrélation maximale entre les vecteurs du dictionnaire, connus du récepteur, et le signal estimé décide la séquence binaire la plus probablement émise.

Un tel système assure un débit $R_b = \frac{F_e}{N}$. Augmenter ce débit par diminution du temps bit N entraîne, comme nous le verrons par la suite, une détérioration du TEB. Il paraît donc intéressant d'introduire un paramètre supplémentaire m permettant de jouer sur le débit sans modifier pour autant le temps bit. Ce paramètre traduit la transmission simultanée d'un m-uplet binaire, sur une durée de N échantillons, au débit $R_s = \frac{mF_e}{N} = mR_b$.

3 Modulations par étalement de spectre

Deux possibilités de modulations par étalement de spectre peuvent être envisagées pour transmettre simultanément un m-uplet binaire (un symbole).

3.1 Modulations de symboles

La première possibilité consiste à augmenter le nombre de vecteurs du dictionnaire, de sorte à établir une relation bijective entre les $M = 2^m$ m-uplets binaires possibles. La théorie des communications numériques [3] montre la dépendance entre la probabilité d'erreur d'une chaîne de communication et la distance Δ entre les vecteurs choisis (au sens de la corrélation). En particulier, pour un dictionnaire de deux vecteurs,

$$\mathcal{P} = Q\left(\sqrt{\frac{\Delta^2}{2\sigma_x^2}}\right) \text{ où } Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty exp^{-\frac{t^2}{2}}\,dt$$

et σ_x est la puissance du signal audio. Dans ce cas, le choix du dictionnaire se porte alors sur deux signaux blancs antipodaux. Au delà pour transmettre simultanément m bits ($m \geq 2$), maximiser cette distance nécessite de choisir des signaux orthogonaux au sens de la corrélation. Ces signaux sont blancs gaussiens (générés par exemple par la fonction "randn" de Matlab) puis orthogonalisés par la procédure d'orthonormalisation de Gram-Schmidt.

3.2 Code Division Multiple Access (CDMA)

La seconde possibilité est propre aux communications numériques : la technique de modulation CDMA [3] permet à plusieurs utilisateurs de partager la même chaîne de transmission. Elle peut être vue comme la transmission d'une information décomposée sur une base de $M = m$ signaux orthogonaux $\mathcal{D} = \{d_1, ..., d_m\}$, i.e.

$$\forall\, n \in [0, N-1],\; v(n) = \sum_{k=1}^{m} a_k d_k(n),$$

où a_k est l'amplitude physique (± 1) associée au k-ième bit du m-uplet binaire $(b_1, ..., b_m)$ transmis. Ces m vecteurs (ou séquences), comme dans le cas de la modulation précédente, sont blanches, gaussiennes et orthogonalisées par Gram-Schmidt.

Chaque vecteur du dictionnaire étant présent dans le signal tatoué et porteur d'information, le détecteur par corrélation est donc remplacé par la recherche de l'information de signe de la corrélation entre le signal modulé estimé et chaque vecteur du dictionnaire. Ces m signes décident du m-uplet binaire vraisemblablement émis. En pratique le m-uplet binaire reçu est obtenu par la relation :

$$\begin{bmatrix} \hat{b}_1 \\ \vdots \\ \hat{b}_M \end{bmatrix} = sgn\left(R_{\mathcal{D}}^{-1}\begin{bmatrix} \hat{v}(0) \\ \vdots \\ \hat{v}(N-1) \end{bmatrix}\right),$$

où $R_{\mathcal{D}}$ est la matrice d'autocovariance du dictionnaire d'émission et $\left[\begin{array}{ccc} \hat{v}(0) & \dots & \hat{v}(N-1) \end{array} \right]^T$ est le signal modulé estimé sur un temps bit.
La minimisation du TEB de transmission nécessite de se prévenir d'éventuelles interférences entre utilisateurs. La théorie des communications numériques s'est donc dotée de séquences pseudo-aléatoires aux propriétés d'intercorrélation très particulières. Parmi elles, les séquences Gold [3] de longueur $N = 2^p - 1$ et aux nombres de $N + 2$ se caractérisent par une intercorrélation égale à -1 (les rendant quasiment orthogonales) et une autocorrélation réduite à 3 valeurs $\{-1, -t(p), t(p) - 2\}$ où

$$t(p) = \begin{cases} 2^{\frac{p+1}{2}} + 1 & si\ p\ impair \\ 2^{\frac{p+2}{2}} + 1 & si\ p\ pair \end{cases} .$$

4 Résultats expérimentaux

4.1 Protocole expérimental

Les performances de ces modulations par étalement de spectre sont évaluées par la donnée du TEB en fonction du débit $R_s = mR_b$. Ce TEB, moyen, est obtenu par tatouage de $\mathcal{B}$ bits d'information d'un échantillon de 5 signaux de musique de style divers (musique classique mono- et pluri-instrumentale, variété), échantillonnés à 44.1 kHz (qualité CD). Le TEB est un estimateur efficace de la probabilité d'erreur de transmission $\mathcal{P}$, pour une précision de $\Delta\mathcal{P} = \sqrt{\frac{TEB}{\mathcal{B}}}$ et un taux de confiance de 70%. Trois constations s'imposent :

- Il paraît illusoire de faire fonctionner le système de tatouage à des débits conduisant à un TEB supérieur à 5%, dans la mesure où le système a vocation de transmettre une information.
- La précision indique qu'il faudrait transmettre 10 millions de bits pour obtenir une précision relative de 10^{-3}, simulation très coûteuse en temps de calcul. Le choix a donc été fait de transmettre $\mathcal{B} = 1000$ bits d'information, bon compromis entre la précision des résultats (0.32% pour un TEB de 1%) et le temps de calcul (12h pour l'ensemble des résultats présentés).
- Les tests du système de tatouage ont montré la forte dépendance des TEB au signal audio à tatouer : pour l'un d'eux, les TEB obtenus sont souvent nettement supérieurs (de l'ordre de deux fois) au TEB moyen. L'origine de ces différences reste encore à clarifier. Néanmoins, les variations des TEB en fonction des débits sont identiques quel que soit le signal utilisé.

Pour s'assurer de la robustesse du système de tatouage à une opération de compression-décompression, le canal est perturbé par un codeur MPEG 1 Layer 1, fonctionnant à un débit de 96 kbits/s.

4.2 Résultats

Les performances des modulations par étalement de spectre envisagées sont présentées figures 2, 3 et 4 : les TEB sont tracés en fonction du débit pour différentes valeurs du nombre de bits transmis simultanément m (se traduisant par un dictionnaire de $M = 2^m$ vecteurs pour les modulations de symboles et $M = m$ vecteurs pour les CDMA). Elles ont en commun la courbe des performances du système initial, obtenues pour une modulation avec deux vecteurs antipodaux, qui constitue notre référence : cette courbe confirme la détérioration du TEB avec la diminution du temps bits. Les traits verticaux représentent la précision des TEB obtenus aux points de mesure.

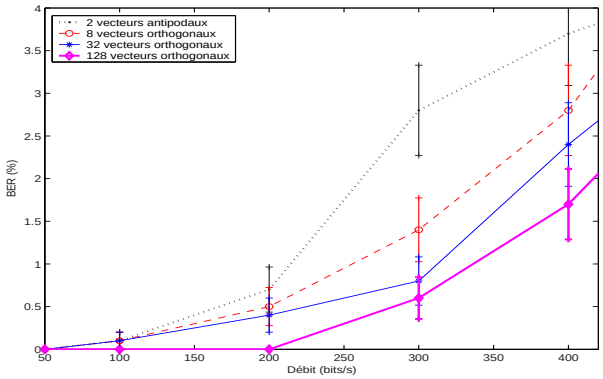

Figure 2 – *TEB pour des modulations de symboles.*

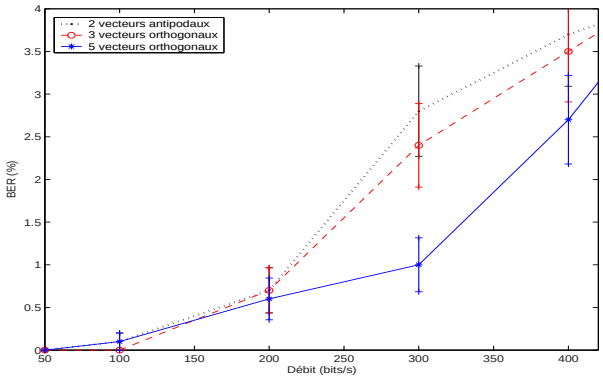

Figure 3 – *TEB pour une modulation CDMA classique.*

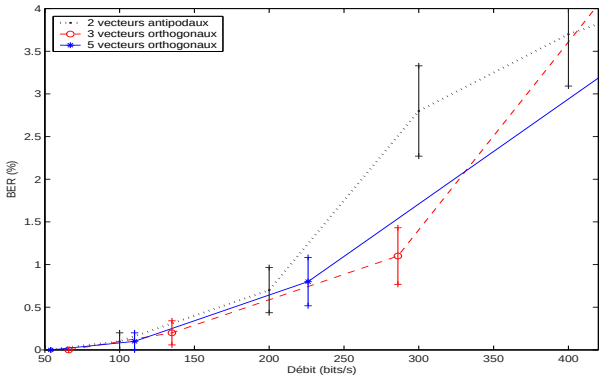

Figure 4 – *TEB dans le cas d'une modulation CDMA utilisant des séquences Gold.*

Concernant les modulations de symboles, figure 2, on constate une nette amélioration du TEB (de l'ordre de 50% pour des débits entre 200 et 400 bits/s par rapport à la référence à partir de $m \geq 5$). Cette amélioration va de pair avec l'augmentation de la taille du dictionnaire. Des constatations similaires peuvent être faites pour les techniques CDMA avec des séquences classiques, figure 3 : le paramètre de taille du dictionnaire joue un rôle important sur cette amélioration (avec notamment une diminution du TEB de 50% à 300 bits/s lorsque $m = 5$). L'influence de ce paramètre est moins visible dans le cas des modulations CDMA avec les séquences Gold, figure 4. Bien que la mise en oeuvre de ces techniques permettent d'améliorer le TEB à 300 bits/s, leurs performances sont inférieures à celles des modulations CDMA classiques avec un dictionnaire de 5 vecteurs. Les séquences Gold ne constituent donc pas le meilleur choix pour un tel système de tatouage. On leur préfèrera les signaux blancs gaussiens.

Le choix de la meilleure modulation se fait donc entre les modulations de symboles et les modulations CDMA classiques, dont les performances à première vue sont similaires lorsqu'un m-uplet avec $m = 5$ est modulé. De plus, l'amélioration du TEB semble conjointe à l'accroissement la taille du dictionnaire : on peut alors se demander si cette tendance se confirme au delà de $m > 5$. La figure 5 vise à répondre à ces deux problèmes : le TEB y est tracé pour deux débits fixés - 300 et 400 bits/s (les plus représentatifs des variations du TEB) - en fonction du paramètre m.

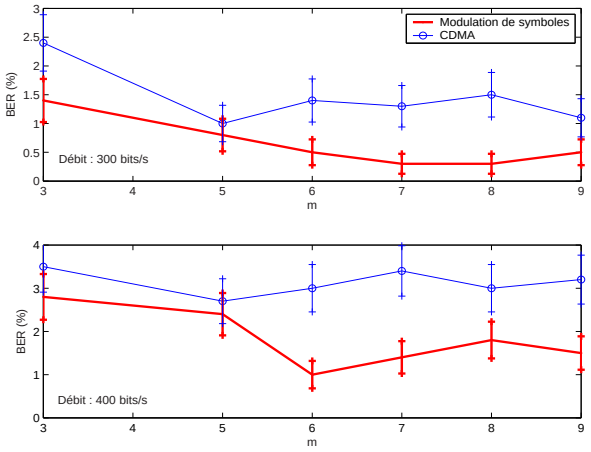

Figure 5 – *TEB en fonction de la taille du dictionnaire à débits fixés.*

Ce graphe confirme l'efficacité des modulations de symboles devant les modulations CDMA classiques quel que soit m (excepté éventuellement pour $m = 5$). Les performances des modulations de symboles ont une tendance décroissante jusqu'à $m = 8$ pour un débit de 300 bits/s et $m = 6$ pour 400 bits/s. Pour $m = 9$ à 300 bits/s, une légère remontée (plus irrégulière à 400 bits/s) semble s'amorcer. Celle-ci aurait peut-être été confirmée si les possibilités de calcul des machines utilisées (notamment l'orthonormalisation par Gram-Schmidt) n'avaient limité les simulations. Cette remontée est sans doute due à une mauvaise estimation de la matrice de covariance du signal modulé pour le filtrage de Wiener étant donné le nombre de vecteurs possibles (pour $m = 9$, $M = 2^m = 512$). Le signal utilisé devrait être allongé, mais un compromis serait alors à faire entre le temps de calcul de Wiener, déjà long, et le TEB obtenu. Conclure à l'existence d'une taille "optimale" pour le dictionnaire est donc difficile.

A ces constations s'ajoute le coût en temps de calcul, primordial dans le cadre de cette application que l'on souhaite être en temps réel. Les temps de réception (filtrage de Wiener et détection binaire) sont équivalents pour les modulations de symboles et les CDMA : leur ordre de grandeur correspond à 66% du traitement complet d'un signal par la chaîne de tatouage.

Pour un tel système, le choix se portera finalement sur une modulation de symboles par étalement de spectre de paramètre $m = 6$ ou 7 (i.e. un dictionnaire de 64 ou 128 vecteurs). Une telle modulation permet d'envisager, dans l'exemple où $m = 7$, la restitution d'une information binaire sans erreur, jusqu'à 300 bits/s pour une transmission idéale (sans perturbation) et jusqu'à 200 bits/s si le canal audio est perturbé par une opération de compression.

5 Conclusion

Différentes techniques de modulations par étalement de spectre ont été implémentées dans un système de tatouage audio à vocation de transmission de données : la modulation de m-uplets binaires associés de manière bijective aux $M = 2^m$ signaux blancs gaussiens d'un dictionnaire ou les techniques CDMA avec différentes séquences, dont les séquences Gold. Toutes ces modulations ont montré leurs apports dans l'amélioration de performances du système en terme de débit et de BER : pour $m \leq 6$, plus la taille du dictionnaire augmente (i.e. plus grand est le nombre de bits envoyés simultanément), meilleur est le TEB à débit équivalent. L'étude se porte en faveur des modulations de symboles, les plus performantes dans cette configuration du système de tatouage : elles permettent d'envisager des transmissions sans erreurs pour des débits inférieurs à 200 bits/s malgré la perturbation du signal audio tatoué par un codeur MPEG, fonctionnant à un débit de 96 kbits/s.

Références

[1] L. de Campos Teixeira Gomes. *Tatouage de signaux audio*. Thèse de doctorat, Université René Descartes - Paris V, Juillet 2002.

[2] C. Baras. Etude de la mise en forme de l'information binaire dans un système de tatouage audio. Rapport technique, ENST, Mémoire de DEA, Septembre 2002.

[3] J.G. Proakis. *Digital communications*. McGraw-Hill, 2001, 4th edition.