Détection par approche turbo codes dans un système de tatouage audio

T. Majoul[1,2] F. Raouafi[1,2] S. Larbi[1] M. Jaidane[1]

[1] Ecole Nationale d'Ingénieurs de Tunis, Unité de recherche Signaux et Systèmes (U2S), Tunis, BP 37, Le Belvédère, 1002, Tunisie.
[2] Institut Supérieur d'Informatique, Ariana, 2, Rue Abou Rayhane Bayrouni, 2080, Tunisie.

U2S@enit.rnu.tn

Résumé

Outre l'aspect de protection de la propriété intellectuelle, les systèmes de tatouage audionumérique deviennent de plus en plus dédiés à des applications multimédia où un flux binaire est transmis d'une façon inaudible et indélébile dans un signal audio original. La récupération du signal émis doit être donc fiable et efficace face aux perturbations licites ou illicites que peut subir le signal audio tatoué. Dans cette optique, nous proposons un système de tatouage fondé sur une détection par approche des turbo codes (décodage itératif). Des résultats expérimentaux explorant la robustesse de la détection sur un ensemble de signaux réels sont proposés pour mettre en évidence les performances de cette stratégie de détection.

Mots clefs

tatouage audio, turbo codes, robustesse face aux attaques.

1 Introduction

Le tatouage audionumérique est vu comme une solution pour de nombreuses applications multimédia où un flux binaire peut être transmis en prenant le signal audio comme support de transmission.
La détection du message inséré fait l'objet de plusieurs travaux de recherche. Citons le système présenté par Malvar [1] qui se base sur la transformation temps/fréquence (Modulated Complex Lapped Transform) du signal audio. Ce système a montré une bonne robustesse à une très grande variété d'attaques mais à des débits de transmission très limités. La méthode élaborée par Cvejic dans [2] présente un système de détection avec des turbo codes mais les détails de la conception n'ont pas été abordés. Ce présent travail se situe dans le contexte de l'amélioration des performances d'une chaîne de tatouage audionumérique par étalement de spectre par une approche turbo codes. Notre travail se base sur un système de référence [3, 4] où la détection est réalisée par un filtre de Wiener.
Pour remédier aux limites de ce système, plusieurs travaux ont été déjà proposés allant des techniques de détection aveugles et semi-aveugles (filtre de Wiener [3], tatouage informé [4, 5]), aux schémas de tatouage avec information adjacente. La procédure de détection présentée dans [3] utilise à la réception le filtre inverse du filtre de mise en forme suivi par un filtre de Wiener. Dans [6], il a été démontré que le bruit équivalent dans une chaîne de tatouage pouvait être modélisé par un bruit gaussien généralisé (GG). Une nouvelle approche de détection avec codage convolutif a montré des résultats prometteurs. En outre, l'application des turbo codes en présence d'un canal à bruit GG a donné de bonnes performances dans un contexte de communications numériques [7]. La mise en oeuvre des turbo codes pour l'amélioration des performances de la détection pour ce schéma de tatouage de référence a été présentée dans [8]. Cette méthode de détection a donné de meilleurs résultats et a montré une bonne robustesse face à la compression MPEG.
Dans ce papier, nous explorons les performances de ce système de tatouage audio en évaluant sa robustesse face à un ensemble de perturbations dont le signal audio tatoué peut être l'objet et en étudiant l'effet de la taille des blocs de tatouage, codés. Nous rappelons tout d'abord, dans la section suivante, les concepts de base du système de tatouage audio testé. Puis, nous présentons le protocole expérimental adopté. Les performances du système sont présentées dans la dernière partie.

2 Système de tatouage audio testé

Nous rappelons ici le système de tatouage audio testé incluant un encodeur et un décodeur itératif [8]. Ce système (figure 1) peut être vu comme une chaîne de communication numérique avec des caractéristiques très particulières. En effet, le bruit, le signal audio original, est à puissance élevée, très coloré et non stationnaire ; en particulier sa densité de probabilité varie d'une fenêtre d'analyse à une autre.

2.1 L'émetteur

Il inclut un encodeur, un modulateur (MOD.) et un filtre de mise en forme. La suite des bits b_i représentant le message binaire à insérer dans le signal audio x_n est introduite dans un module de codage canal pour former une autre suite binaire codée de L symboles a_k (chaque symbole a_k est le groupement de l bits codés). La séquence des symboles a_k est ensuite modulée à l'aide d'un dictionnaire D

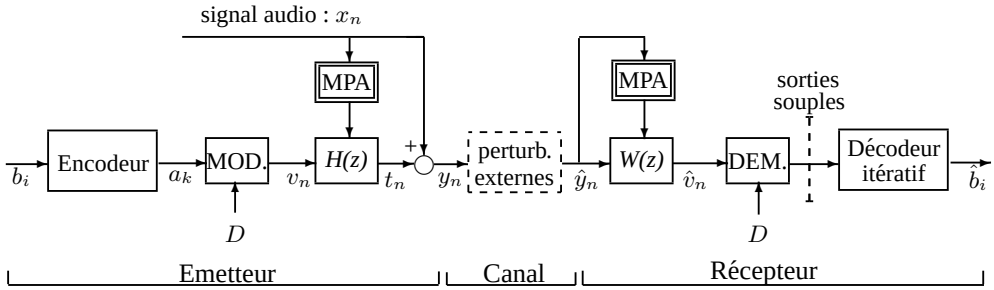

Figure 1 – *Système de tatouage audionumérique proposé, incluant un encodeur et un décodeur itératif [8].*

contenant M vecteurs d_k blancs, gaussiens et orthogonaux deux à deux ($M = 2^l$ où l le nombre de bits par symbole). A chaque symbole a_k est associé un vecteur d_k du dictionnaire D. Ainsi, la concaténation de la suite des L vecteurs d_k correspondant à la signature à cacher dans le signal audio forme le signal modulé v_n. Ce dernier est ensuite mis en forme spectralement par un filtre $H(f)$ dont la réponse impulsionnelle est obtenue à partir d'un modèle psychoacoustique (MPA)[1] donnant le seuil de masquage caractérisant la contrainte d'inaudibilité et actualisée dans chaque fenêtre d'analyse de longueur N. Une addition dans le domaine temporel du signal audio x_n et du signal mis en forme t_n permet d'obtenir le signal audio tatoué y_n.

2.2 Le canal

Il est à l'origine de la détérioration du signal audio tatoué y_n suite, par exemple, à une manipulation licite telle que la compression MPEG ou une attaque pirate telle que le filtrage passe-tout. Le signal résultant $\hat{y}_n$ est une version bruitée du signal tatoué y_n.

2.3 Le récepteur

Il comporte un filtre de Wiener, un démodulateur (DEM.) et un décodeur itératif. Le signal $\hat{y}_n$ est traité d'abord par un filtre de Wiener $W(z)$ minimisant de manière implicite l'erreur quadratique moyenne $E[(v_n - \hat{v}_n)^2]$. Ce caractère aveugle de la minimisation, sans connaissance explicite de v_n, est permi par la connaissance à priori de la corrélation des éléments du dictionnaire d'émission et la connaissance à priori du filtre de mise en forme H (supposé proche de celui issu du signal reçu $\hat{y}_n$).
Suivant les caractéristiques du signal audio dans la fenêtre d'analyse courante, chaque vecteur d_k émis du dictionnaire sera bruité et on aura donc à la réception (après les étapes de mise en forme par le filtre H, les perturbations éventuelles sur le signal tatoué et le filtrage de Wiener) un vecteur $\hat{v}_k$. Ce vecteur sera donc déphasé d'un certain angle ϕ_k par rapport au vecteur initialement transmis.
Les valeurs de déphasage peuvent-être déduites à partir des valeurs des intercorrélations des vecteurs $\hat{v}_k$ constituant le signal $\hat{v}_n$ reçu et les vecteurs d_k du dictionnaire de modulation. Celles ci sont nécessaires pour l'étape de décodage (voir annexe).

[1]Le MPA utilisé est une version modifiée du modèle n°1 de MPEG adaptée au contexte de tatouage.

3 Protocole expérimental

Les performances du système de détection proposé sont évaluées en terme de taux d'erreurs binaire (TEB) moyens résultant de simulations Monte Carlo de la détection de messages binaires aléatoires sur 5 signaux test de styles différents (musique classique, instrumentale et voix chantée) échantillonnés à 44.1 kHz.
La probabilité d'erreur de la détection peut être estimée par le TEB avec une précision $p = \sqrt{TEB(1-TEB)/B}$ où B est le nombre total de bits émis [5]. Dans notre cas, le nombre de bits total émis est $B = 170000$ bits, ce qui permet d'avoir des résultats avec une précision $p = 0.024\%$ pour un TEB de 1%.

• **Systèmes de détection évalués :** trois techniques de détection du tatouage sont évaluées :
- filtrage de Wiener seul (système de référence),
- filtrage de Wiener + codeur convolutif,
- filtrage de Wiener + turbo code à 3 et 5 itérations.

• **Perturbations considérées :** les performances de la détection sont évaluées dans le cas d'un canal sans perturbation et en présence de perturbations non désynchronisantes. Le tableau 1 récapitule les situations de canal envisagées.

CANAL	Description
C1	sans perturbation
C2	ajout de bruit
C3	filtrage passe-tout
C4	MPEG (96 kbits/s)
C5	MPEG (64 kbits/s)

Tableau 1 – *Description des situations de canal considérées.*

C1 : un canal sans perturbation.
C2 : un ajout de bruit blanc à un rapport signal à bruit de 30 dB.
C3 : un filtrage passe-tout d'ordre 2 avec un filtre dont les zéros sont donnés par $[-8; 5]$.
C4 et C5 : une opération de compression/reconstruction

MPEG[2] à des débits respectifs de 96 et 64 kbits/s.

• **Paramètres du turbo code utilisé :** l'encodeur utilisé est formé de deux codeurs récursifs systématiques en parallèle (ayant le même polynôme générateur $(37, 25)$) qui sont utilisés également pour le codeur convolutif. Le rendement du codage est $R = 1/2$. Le décodage itératif est basé sur l'algorithme du maximum à posteriori (MAP) [10].

4 Résultats expérimentaux

Les figures 2 et 3 montrent les TEB obtenus pour un dictionnaire de modulation composé de 2 vecteurs orthogonaux. Le débit de transmission est 43 bits/s dans ce cas (en tenant compte du rendement du codage $R = 1/2$).
La figure 2 présente les performances du système de référence et celles du système proposé (codeur convolutif et turbo code à 3 et à 5 itérations) dans le cas de blocs de données binaires représentant le tatouage de taille 680 bits. La figure 3 met en évidence l'amélioration des performances avec des blocs de taille 1700 bits.

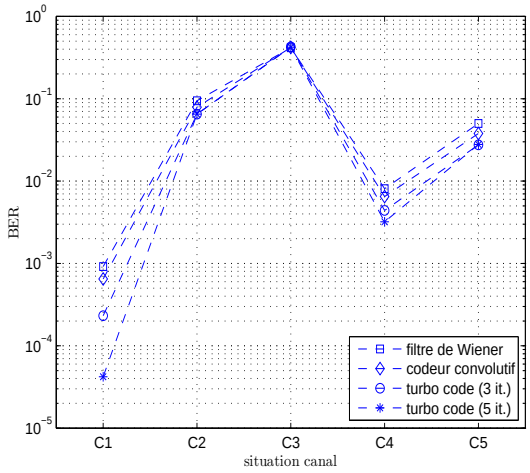

Figure 2 – *TEB obtenus selon différentes méthodes de détection pour différentes situations canal (Tableau 1) dans le cas de données binaires représentant le tatouage en blocs de* 680 *bits. Débit utile* = 43 *bits/s.*

4.1 Performances

Les résultats obtenus (figure 2) montrent une amélioration sensible de la détection.
C1 : dans le cas d'un canal sans perturbation, on peut atteindre des TEB de l'ordre de 10^{-4} avec un turbo code à 3 itérations et des TEB de l'ordre de 10^{-5} à la 5ème itération alors qu'avec un filtrage de Wiener le TEB est de l'ordre de 10^{-3}.
C2 : concernant le canal avec un ajout de bruit à 30 dB, on remarque aussi une légère amélioration avec le codage convolutif et avec un turbo code.

[2]Réalisée par le codeur LAME

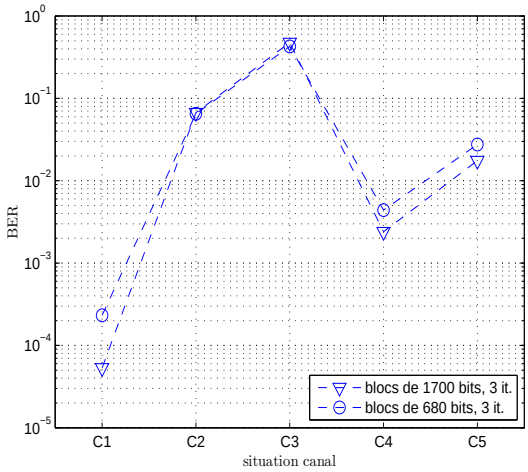

Figure 3 – *TEB obtenus selon une détection avec un turbo code à* 3 *itérations pour différentes situations canal (Tableau 1) pour des données binaires représentant le tatouage en blocs de* 1700 *bits et* 680 *bits. Débit utile* = 43 *bits/s.*

C3 et C4 : dans le cas d'une compression MPEG, on constate également une amélioration du TEB par rapport à une détection par un filtre de Wiener seul.
C5 : par contre, le codage canal n'apporte pas d'amélioration dans le cas d'un filtrage passe-tout et la dégradation est aussi importante pour les trois types de détection. En effet, pour ce type de perturbation le signal tatoué subit un déphasage sans pour autant qu'il soit altéré en amplitude. Une égalisation adéquate peut donc être utilisée pour remédier à l'effet du filtre passe-tout.

4.2 Effet de la taille des blocs de données de tatouage codées

Concernant l'effet de la taille des blocs codés, il est clair d'après les résultats (figure 3) que les performances sont meilleures dans le cas de blocs de taille 1700 bits. En effet, dans le cas d'un canal sans perturbation, on passe d'un TEB de 2.10^{-4} dans le cas avec des blocs de tatouage codés de 680 bits à un TEB de 5.10^{-5} avec des blocs de 1700 bits et ce à la 3ème itération. Une légère amélioration est aussi observée dans le cas d'une compression MPEG. Cela s'explique par le fait que les valeurs des probabilités conditionnelles des bits reçus sont plus précises puisque l'estimation de la variance de la phase se fait sur un nombre de réalisations plus grand.

5 Conclusion

Ce papier présente un système de tatouage audio amélioré par l'insertion d'un encodeur et d'un décodeur itératif. Les performances du système proposé ont démontré une nette amélioration en terme de TEB surtout dans le cas d'un canal en absence de perturbations et dans le cas d'une compression MPEG. Par ailleurs, l'augmentation de la taille

des blocs des données binaires représentant le message à insérer a aussi amélioré la détection. Une modélisation de la distribution du déphasage devrait améliorer davantage ces performances. Ces résultats montrent ainsi que l'utilisation des turbo codes dans le système de tatouage améliore sensiblement les performances de détection au prix d'un débit de transmission multiplié par un facteur R, où R est le rendement de codage.
L'augmentation du débit de transmission, tout en utilisant des turbo codes, pour de mêmes performances de détection passe par une optimisation à la fois au niveau du dictionnaire d'émission (longueur, taille des vecteurs,...) et au niveau du récepteur : critère d'optimisation du filtre de réception, estimation des paramètres définissant les entrées souples du décodeur itératif,...

ANNEXE : Calcul des probabilités conditionnelles des bits reçus

L'étape de décodage suggère une connaissance des valeurs des probabilités conditionnelles des bits reçus $P(\hat{b}_i|b_i)$. A partir des valeurs de corrélation entre les vecteurs $\hat{v}_k$ constituant le signal estimé $\hat{v}_n$ et les vecteurs d_k du dictionnaire de modulation D, on calcule la probabilité conditionnelle de chaque vecteur reçu $\hat{v}_k$. Cette valeur est déduite à partir de l'angle ϕ_k (déphasage) entre le vecteur reçu $\hat{v}_k$ et le vecteur d_k du dictionnaire initialement transmis selon l'équation [8] :

$$P(\hat{v}_k|d_k) = f(\phi_k) \tag{1}$$

où f est la loi qui décrit la distribution de la phase dans le bloc de données de tatouage codées reçues. Si l'on suppose que ce déphasage suit une loi gaussienne centrée, on peut écrire :

$$f(\phi_k) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[\frac{-(\phi_k)^2}{2\sigma^2}\right] \tag{2}$$

où σ représente l'écart type de la distribution de ϕ_k sur le bloc de données représentées par les vecteurs $\hat{v}_k$ reçus. Cette valeur est calculée et actualisée pour chaque bloc. Ainsi, on peut estimer la probabilité conditionnelle $P(\hat{b}_i|b_i)$ de chaque bit reçu $\hat{b}_i$.
Par exemple, dans le cas d'un dictionnaire de modulation à deux vecteurs d_0 et d_1, chaque symbole a_k représente un bit envoyé b_i (0 ou 1). Pour un bit reçu $\hat{b}_i$, on peut alors écrire :

$$P(\hat{b}_i \mid b_i = k) = P(\hat{v}_k \mid d_k) = f(\phi_k) \tag{3}$$

où $k = (0, 1)$ et ϕ_k est l'angle formé par le vecteur reçu $\hat{v}_k$ et le vecteur du dictionnaire d_k associé au symbole a_k transmis.

Références

[1] D. Kirovski et S. Malvar, "Robust Spread-Spectrum Audio Watermarking", Proc. Int. Conf. Acoust., Speech and Signal Processing, 2001.

[2] N. Cvejic, D. Tujkovic and T. Seppanen, "Increasing Robustness of an Audio Watermark Using Turbo Codes". Proc. 2003 IEEE. International Conference on Multimedia and Expo., Baltimore, MD, 1217-1220.

[3] S. Larbi, M. Jaïdane et N. Moreau, "A New Wiener Filtering Based Detection Scheme for Time Domain Perceptual Audio Watermarking," IEEE Int. Conf. On Acoustics, Speech, and Signal Processing, Montréal 2004.

[4] C. Baras, N. Moreau et P. Dymarski, "Controlling the Inaudibility and Maximizing the Robustness in an Audio Annotation Watermarking System", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 5, Septembre 2006.

[5] C. Baras et N. Moreau : Modulation CDMA informée dans un système de tatouage audio. Dans COmpression et REprésentation des Signaux Audiovisuels (CORESA), Rennes, France, Novembre 2005.

[6] S. Saied, S. Larbi, R. Hamza, L. Belhadj Slama et M. Jaidane, "Calcul de la Probabilité d'Erreur pour une Chaîne de Tatouage Audionumérique à Bruit Non Gaussien", Gretsi 2003.

[7] Xiaoling, H., and Nam, P., "Turbo Decoders Which Adapt to Noise Distribution Mismatch". *IEEE Communications Letters*, 1998, Vol. 2, No. 12, pp. 321-323

[8] T. Majoul, F. Raouafi et M. Jaidane, "Audio Data Hiding : Improving Detection Using Turbo Codes", 30th AES conference on intelligent audio environments, Saariselka, Finland, 2007.

[9] J. Proakis. Digital Communications. McGraw-Hill, 4ème édition, 2001.

[10] C. Berrou, A. Glavieux and Thitimajshima, "Near Shannon limit error-correction coding : turbo codes," Proc. IEEE Int. Conf. on Communications, Geneva, Switzerland, 1993, pp. 1064-1070.