Système de tatouage audio en boucle fermée

Cléo BARAS[1], Przemyslaw DYMARSKI[2], Nicolas MOREAU[1]

[1]GET - Télécom Paris - Département TSI
46 rue Barrault, 75013 Paris Cédex 13, France

[2]Université Technique de Varsovie
15/19 rue Nowowiejska, 00-665 Varsovie, Pologne
baras@tsi.enst.fr, dymarski@tele.pw.edu.pl,moreau@tsi.enst.fr

Résumé – Un système de tatouage peut être vu comme une chaîne de communication particulière : le canal audio est susceptible de transmettre une information binaire, inaudible et indélébile. L'exploitation de ce canal requière un couple débit-fiabilité de transmission le plus élevé possible. Dans cet objectif, un schéma de réception s'approchant du récepteur optimal défini pour une configuration AWGN est proposé. Une structure en boucle fermée est ensuite envisagée, afin d'exploiter la connaissance du signal support à l'émetteur. Des résultats expérimentaux montrent qu'un tel système fournit un taux d'erreur binaire de transmission inférieur à 5% à un débit de 350 bits/s et garantit une robustesse à la compression MPEG et à la conversion analogique/numérique.

Abstract – Audio watermarking scheme can be designed to use audio signal as a transmission channel for binary information. In this context, transmission rate and reliability criteria are major issues. We expound to new approaches : an adapted implementation of the optimum receiver for an AWGN communication channel and a closed loop watermarking scheme introducing a local copy of this reception process at the embedder. The a priori knowledge of the audio signal is there taken into account during the embedding strategy. Experimental results reveal the efficiency of this informed embedding scheme.

1 Introduction

Initialement développé comme une solution potentielle à la protection des droits de propriétés intellectuelles, le tatouage s'est depuis étendu à de nouveaux domaines d'application, pour lesquels un système de tatouage est vu comme une chaîne de communication : le signal audio (signal de musique pour une qualité CD) est susceptible de porter une information binaire, inaudible et indélébile. L'exploitation des possibilités de transmission de ce canal, très particulier, est l'objet de nos travaux. L'information binaire insérée se caractérise par un débit et une fiabilité de détection, que l'on souhaite les plus élevés possible. Un système de tatouage doit de fait satisfaire à des objectifs de performances, définies en terme de débit et de taux d'erreur binaire (TEB) et de robustesse aux distorsions classiques, telles des opérations de compression-reconstruction ou de conversions analogique/numérique.
Une telle chaîne de communication bénéficie également à l'émetteur de la connaissance du signal support, dans lequel est "noyée" l'information tatouée [1]. Costa [2] montre que la capacité du canal peut être rendue indépendante du signal audio. Pour atteindre des performances optimales, l'émetteur doit s'adapter au signal support plutôt que le récepteur d'en annuler les effets. Notre attention s'est donc portée sur la stratégie d'insertion [3], visant à choisir un tatouage adapté, conciliant distorsion perceptuelle et fiabilité de détection. On parle de tatouage informé.
Cet article présente dans une première partie un système de tatouage de référence, base de notre travail. Il expose dans une seconde partie un schéma de réception, s'approchant du récepteur optimal défini pour un canal AWGN. L'insertion d'une copie locale de ce récepteur à l'émetteur est proposée dans une troisième partie, conduisant à une structure en boucle fermée du système de tatouage. Une stratégie d'insertion maximisant la robustesse du tatouage aux bruits introduits par les perturbations du canal est alors proposée : elle conduit à une adaptation "rétroactive" du signal de tatouage au signal audio, permettant de minimiser les erreurs de transmissions. Les performances expérimentales de ces contributions sont présentées dans une quatrième partie.

2 Principe du système de tatouage de référence

Le système de tatouage de référence, développé dans [6, 7], pour des signaux audio-fréquences échantillonnés à 44.1kHz, est structuré sous la forme d'une chaîne de communication dont le diagramme fonctionnel est donné figure 1.

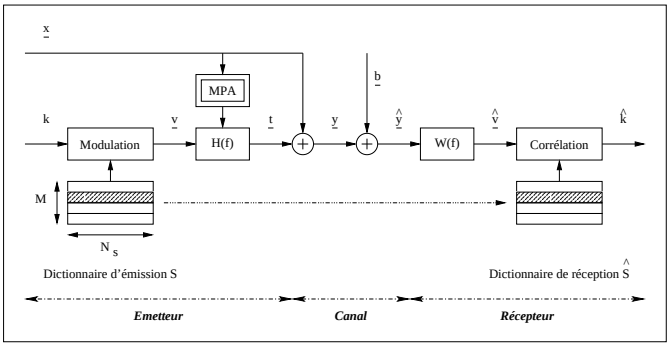

FIG. 1: *Diagramme fonctionnel du système de tatouage de référence.*

Le signal modulé $\underline{v}$ est obtenu par concaténation d'une suite de vecteurs $\underline{s}_m$, la suite des indices m étant caractéristique des symboles transmis (N_{bs}-uplets binaires). Chaque vecteur $\underline{s}_m$ est choisi parmi un dictionnaire d'émission S, constitué de $M = 2^{N_{bs}}$ signaux blancs, gaussiens, de durée N_s, de puissance unité, orthogonaux. Le signal modulé est mis en forme spectralement par un filtre $H(f)$, dont la réponse en fréquence est actualisée toutes les 10 ms environ. Ce filtrage a pour objectif d'adapter la densité spectrale de puissance de $\underline{v}$ à un seuil de masquage (issu d'un modèle psychoacoustique (MPA)), limite fréquentielle caractérisant la contrainte d'inaudibilité. Le signal tatoué $\underline{y}$ est finalement obtenu par sommation temporelle entre le signal audio $\underline{x}$ et le signal mis en forme $\underline{t}$.
Deux formes de perturbation $\underline{b}$ du signal audio tatoué $\underline{y}$ sont envisagées :

- une opération de compression-reconstruction, réalisée par un codeur MPEG. En provoquant une forte distorsion des hautes fréquences, elle impose aux vecteurs du dictionnaire S d'être limités à la bande de fréquence $[0, F_c]$. On choisit $F_c = 11kHz$.
- une opération désynchronisante réalisée par la transmission via une ligne analogique de $\underline{y}$ entre deux PCs. Outre le décalage temporel introduit, la désynchronisation naît de la différence des fréquences d'échantillonnage entre le PC émetteur, qui génère le tatouage et le PC récepteur, qui détecte l'information transmise. Elle nécessite l'ajout dans $\underline{v}$ d'un vecteur de synchronisation régulier précédant chaque vecteur $\underline{s}_m$ porteur d'information et réalisant au récepteur la synchronisation de rythme.

Le signal reçu $\hat{\underline{y}} = \underline{y} + \underline{b}$ est soumis à un filtrage de Wiener qui blanchit le signal audio et estime le signal modulé $\hat{\underline{v}}$. La détection de l'information est finalement réalisée par calcul des intercorrélations entre le signal $\hat{\underline{v}}$ et les vecteurs d'un dictionnaire de réception, $\hat{\mathcal{S}} = \{\hat{\underline{s}}_m\}_{m=0..M-1}$, identique au dictionnaire d'émission. Le vecteur maximisant la corrélation fournit le symbole reçu sur chaque fenêtre d'analyse de durée N_s.

3 Vers un schéma de réception optimal

La théorie des communications numériques [4] montre que le détecteur par corrélation, utilisé par le système de référence, est optimal au sens du minimum de la probabilité d'erreur de transmission dans le cas d'un canal AWGN. Une étape de blanchiment du signal audio coloré, proposée dans [5], permet d'approcher une telle configuration. En effet, remplaçons le filtre de Wiener $W(f)$ par le filtre blanchissant $G(f)$ de $\underline{x}$. Le signal reçu s'écrit $\hat{\underline{v}} = \underline{r} + \underline{w}$, où $\underline{w}$ est le signal utile $\underline{t}$ filtré par $G(f)$, contenant l'information de tatouage et $\underline{r}$, le signal audio filtré par $G(f)$ *i.e.* un bruit additif relativement blanc. Le détecteur optimal est alors un détecteur par corrélation, qui décompose le signal reçu sur les vecteurs d'un dictionnaire de réception $\hat{\mathcal{S}}$. Ce dictionnaire est constitué de l'ensemble des formes prises par l'information tatouée $\underline{w}$, c'est à dire les M vecteurs du dictionnaire d'émission filtrés par $H(f)G(f)$. Le signal audio n'étant pas disponible lors de la procédure de détection, les filtres $H(f)$ et $G(f)$ sont estimés à partir du signal audio tatoué $\hat{\underline{y}}$. Le schéma global d'un tel système est donné figure 2.

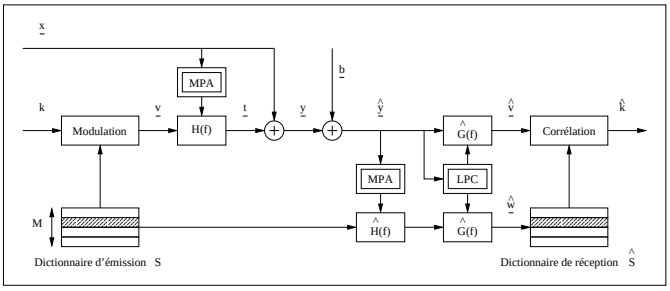

FIG. 2: *Schéma du système de tatouage, dont la réception est basée sur un filtre blanchissant.*

4 Tatouage en boucle fermée

4.1 Position du problème

Étant donnée la configuration du système de tatouage précédant, la condition nécessaire et suffisante d'une transmission sans erreur de l'information k s'écrit :

$$\forall m \neq k, (\hat{\underline{w}} + \hat{\underline{r}} + \hat{\underline{b}})^t \hat{\underline{s}}_k > (\hat{\underline{w}} + \hat{\underline{r}} + \hat{\underline{b}})^t \hat{\underline{s}}_m, \qquad (1)$$

où $\hat{\underline{r}}$, $\hat{\underline{w}}$ et $\hat{\underline{b}}$ sont respectivement le signal audio, le signal de tatouage et le bruit introduit par la perturbation du canal, filtrés par $\hat{G}(f)$. Dans l'hypothèse où $H(f)$, $G(f)$ obtenus à l'émetteur et $\hat{H}(f)$ et $\hat{G}(f)$ sont peu différents, l'insertion d'une copie locale du récepteur à l'émetteur donne accès à une estimation $\tilde{\underline{r}}$, $\tilde{\underline{w}}$ et $\tilde{\underline{S}} = \{\tilde{\underline{s}}_m\}_{m=0..M-1}$ des paramètres de réception $\hat{\underline{r}}$, $\hat{\underline{w}}$ et $\hat{\underline{S}}$. Le système de tatouage présente alors une structure en *boucle fermée*, donnée figure 3. En l'absence de bruit, cette structure permet de prédire le résultat de la détection et de déterminer en conséquence un tatouage optimal.
Les recherches menées dans le sens du tatouage informé convergent vers un découpage de l'espace signal en deux régions d'intérêt [3]:

- *la région de distorsion acceptable*, ensemble des signaux perceptuellement proches du signal audio original. Elle est définie par la contrainte de puissance $\frac{||\underline{v}||^2}{N_s} \leq 1$ imposée au signal modulé pour garantir l'inaudibilité de $\underline{t}$ après sa mise en forme par $H(f)$.
- *la région de détection du tatouage*, ensemble des signaux tatoués pour lesquels l'information k est correctement détectée, satisfaisant à l'équation 1. Dans cette équation seul $\hat{\underline{b}}$ ne peut être estimé par la copie locale du récepteur. Les corrélations $(\hat{\underline{w}}+\hat{\underline{r}})^t(\hat{\underline{s}}_k-\hat{\underline{s}}_m), \forall m \neq k$ doivent alors être maximisées pour assurer la robustesse du tatouage au bruit. Cette robustesse peut s'exprimer en introduisant un seuil σ, que l'on souhaite le plus élevé possible, de manière analogue à la stratégie proposée par Miller dans [3] appelée *Maximising robustness*. Le tatouage vérifie alors :

$$\forall m \neq k, (\tilde{\underline{w}} + \tilde{\underline{r}})^t(\tilde{\underline{s}}_k - \tilde{\underline{s}}_m) > \sigma. \qquad (2)$$

Ce découpage indique que le tatouage recherché doit donc être choisi dans l'intersection de ces deux régions. Pour être optimal, il doit également présenter une puissance maximale, critère nécessaire à sa détection. Sa puissance optimale, limitée par la région de distorsion acceptable, sera donc égale à 1.

4.2 Existence d'un tatouage solution

L'intersection des deux régions d'intérêt pour un paramètre de robustesse fixé pouvant être vide, il est nécessaire de déterminer l'existence d'un tatouage solution inaudible. Une solution existe si le tatouage de plus faible puissance assurant l'équation 2 est dans la région de distorsion acceptable. Le problème d'optimisation sous contraintes suivant doit alors être résolu :

$$\text{Trouver } \underline{v}_\sigma = \arg_{\underline{v}} \min ||\underline{v}||^2 \quad \text{tel que } \begin{cases} \forall m \neq k, (\underline{\tilde{w}} + \underline{\tilde{r}})^t (\underline{\tilde{s}}_k - \underline{\tilde{s}}_m) > \sigma \\ \frac{||\underline{v}||^2}{N_s} \leq 1 \end{cases} . \quad (3)$$

Si $||\underline{v}_\sigma||^2 > N_s$, aucun signal $\underline{v}$ solution de l'équation 2 n'est satisfaisant puisqu'il ne respecte pas la contrainte d'inaudibilité. Sinon $\underline{v}_\sigma$, dont la puissance n'est pas nécessairement optimale, permettra une transmission sans erreur de k.
La solution $\underline{v}_\sigma$ est recherchée sous la forme d'une combinaison linéaire des vecteurs du dictionnaire, définissant l'espace de tatouage, dont les coefficients sont obtenus par la méthode d'Uzawa, décrite dans [8].

4.3 Recherche du tatouage optimal

Le tatouage optimal recherché conjugue puissance maximale (i.e. 1) et robustesse maximale σ_{opt}, précaution contre le bruit. Or on constate que σ augmente lorsque la fonction $||\underline{v}_\sigma||^2$ augmente. En faisant tendre $||\underline{v}_\sigma||^2$ vers 1, on converge alors vers un tatouage de robustesse au bruit maximale, c'est à dire le tatouage optimal, tel que nous l'avons défini.
La convergence est obtenue en faisant varier σ de manière dichotomique après analyse de la puissance $\frac{||\underline{v}_\sigma||^2}{N_s}$ associée.

5 Performances

5.1 Protocole expérimental

Les performances des contributions proposées sont évaluées par la donnée du TEB en fonction du débit $R = \frac{N_{bs} F_e}{N_s}$. Le TEB, moyen, est obtenu par tatouage de B bits d'information d'un échantillon de 5 signaux de musique. Ces signaux, échantillonnés à $F_e = 44.1kHz$, sont de style divers (musique classique mono- et pluri-instrumentale, variété) et de durée $\frac{B}{R}$ secondes. Des simulations préliminaires montrent la forte dépendance du TEB et du signal audio tatoué. Néanmoins, les variations du TEB en fonction du débit sont identiques quels que soient les signaux audio choisis.
Le TEB est un estimateur efficace de la probabilité d'erreur de transmission P de la chaîne de tatouage, pour une précision de $\sqrt{\frac{TEB(1-TEB)}{B}}$ et un taux de confiance de 70%. Cette 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 B = 1000 bits d'information, bon compromis entre la précision des résultats (0.32% pour un TEB de 1%) et le temps de calcul (24h pour l'ensemble des résultats présentés) d'un programme écrit en language Matlab©version 6.1 et simulé sur un Pentium 4, 1.80 GHz.

Les performances sont présentées pour trois types de canal : un canal sans perturbation, une opération de compression réalisée par un codeur MPEG 1 Layer 1 fonctionnant à 96 kbits/s et une opération désynchronisante réalisée par transmission analogique, pour laquelle le débit est limité par les performances de la méthode de synchronisation.

5.2 Résultats

Les performances du système en terme de TEB pour un dictionnaire de 4 vecteurs orthogonaux ($N_{bs} = 2$) sont données figure 4 pour un canal *sans perturbation*, figure 5 pour un canal avec *compression MPEG* et figure 6 dans le cas d'une *conversion analogique*. Ces figures présentent chacune trois courbes : la première donne les performances du système de référence, schéma en boucle ouverte utilisant un filtre de Wiener ; la seconde présente les TEBs obtenus lorsque la réception exploite le filtre blanchissant du signal audio tatoué ; la dernière fait état des performances du système structuré en boucle fermée.
Dans le cas d'une structure en boucle ouverte, ces courbes montrent que la réception basée sur le filtre blanchissant du signal audio tatoué présente des performances équivalentes au schéma de réception de référence. En effet, étant donnée la précision des résultats, figurée aux points de mesure par des traits verticaux, les TEBs obtenus pour ces deux méthodes sont du même ordre de grandeur, quel que soit le canal considéré. Pour des applications temps-réel, le temps de calcul peut être un facteur déterminant dans le choix d'un schéma de réception. La table 1 donne le rapport entre le temps de détection de l'information et la durée du signal tatoué en fonction du débit. Pour ce dictionnaire, de petite taille, les temps de calcul des deux méthodes de réception sont donc équivalents pour des débits élevés (à partir de 500 bits/s).
Ces courbes confirment également l'apport de la prise en compte du signal hôte à l'émetteur sur les performances du système. Une très nette amélioration des TEBs est constatée dans le cas d'un canal sans perturbation. Quel que soit le débit, les TEBs sont ainsi diminués de plus 50% par rapport à ceux d'un système en boucle ouverte. Ainsi, une transmission avec un TEB inférieur à 2% est envisageable jusqu'à 600 bits/s, alors qu'elle est limitée à 300 bits/s pour un schéma en boucle ouverte. Cette amélioration est confortée par les courbes obtenues dans le cas d'un canal bruité. La robustesse du système à une compression MPEG ou à une conversion analogique/numérique est en effet améliorée de 45%. Un TEB inférieur à 4% est obtenu jusqu'à 600 bits/s (contre 400 bits/s pour le système en boucle ouverte) en cas de compression MPEG. Cette même valeur est obtenue jusqu'à 300 bits/s (contre 200 bits/s) si une transmission analogique du signal tatoué est effectuée.

TAB. 1: Rapports temps de simulation - temps réel en fonction de la méthode de détection et du débit

Débit \ Méthode	Wiener	Filtre Blanchissant
200 bits/s	4.2	8.8
400 bits/s	7.1	10.2
800 bits/s	13.9	12.2

6 Conclusion

Dans le but d'exploiter au mieux les possibilités du canal de transmission qu'est le signal de musique, deux approches ont été proposées : un schéma de réception, approchant le récepteur optimal défini pour les canaux AWGN et une structure en boucle fermée du système de tatouage, permettant l'élaboration d'une stratégie de tatouage adaptée au signal hôte et robuste au bruit. Des résultats expérimentaux montrent la pertinence de cette structure qui permet d'améliorer les TEBs de 50% en moyenne par rapport à un schéma classique en boucle ouverte. Cette amélioration est effective dans le cas d'un canal perturbé par une compression MPEG ou par une transmission analogique du signal audio tatoué.
La stratégie d'insertion détermine un tatouage optimal, obtenu à partir d'un dictionnaire donné. Des études prometteuses sont actuellement menées pour choisir un dictionnaire adapté au signal hôte.

Références

[1] I. Cox, M. Miller et A. McKellips. *Watermarking as communications with side information*. Proceedings of the IEEE, 1999.

[2] M. Costa. *Writing on dirty paper*. IEEE Transaction on Information Theory, 1983.

[3] M. Miller, I. Cox et J. Bloom. *Informed embedding : exploiting image and detectot information during watermark insertion*, International Conference on Image Processing, 2000.

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

[5] H. Kim. *Stochastic model based audio watermark and whitening filter for improved detection*. Proceedings of the International Conference on Acoustic, Speech, Signal Processing, 2000.

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

[7] C. Baras et N. Moreau. *Modulations par étalement de spectre dans un système de tatouage audio*. CORESA, 2003.

[8] P. G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Edition Dunod, 1998.

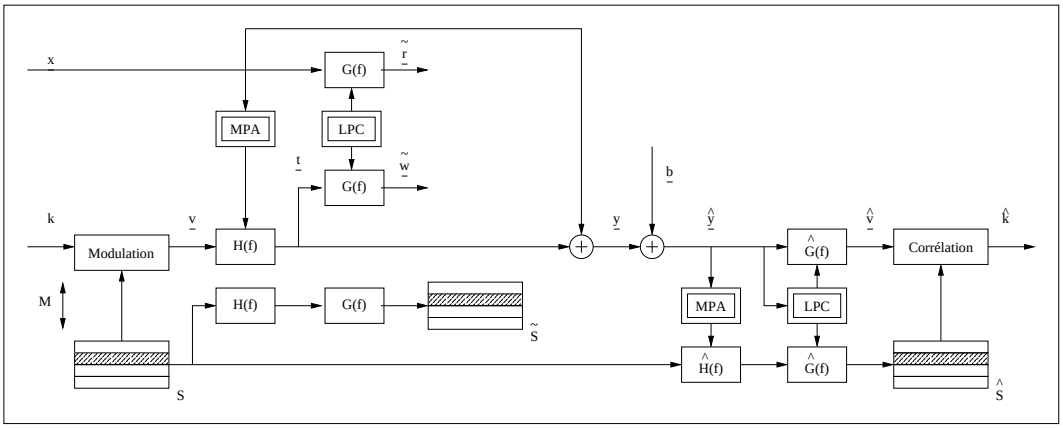

FIG. 3: *Schéma de tatouage en boucle fermée.*

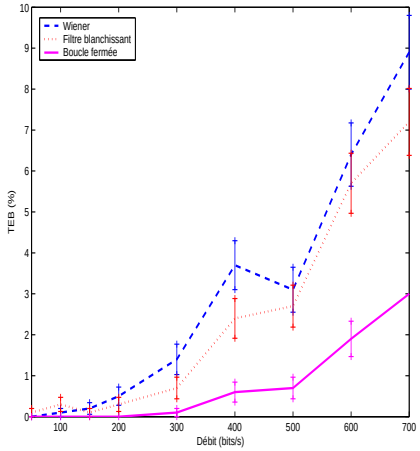

FIG. 4: *Performances des contributions dans le cas d'un canal sans perturbation.*

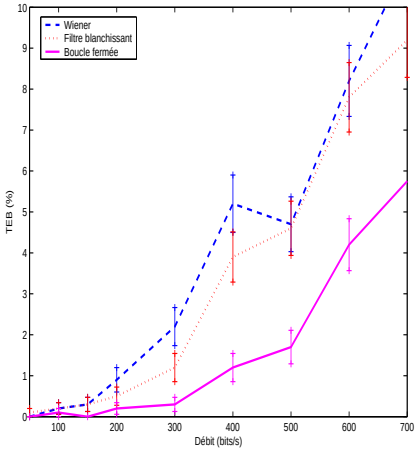

FIG. 5: *Performances des contributions pour un canal perturbé par une compression MPEG.*

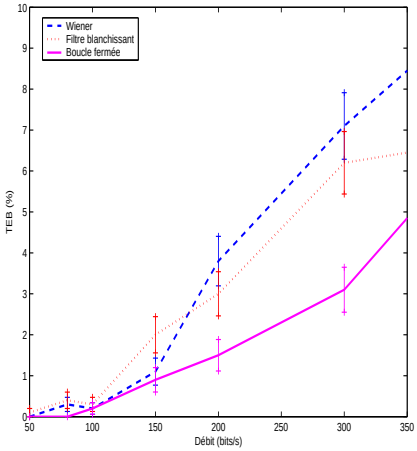

FIG. 6: *Performances des contributions pour une transmission analogique du signal tatoué.*