

Implémentation d'un système de tatouage pour la transmission de données

Alejandro LoboGuerrero, Joël Liénard, Patrick Bas,
Laboratoire d'Images et Signaux,
38402 St Martin d'Hères cedex

mail: alobogue@lis.inpg.fr, joel.lienard@lis.inpg.fr, patrick.bas@lis.inpg.fr http: www.lis.inpg.fr

ABSTRACT

Audio watermarking is a method that allows the insertion of an imperceptible mark on an audio data set. Although the watermarking is often used to guarantee copyrights, it can also be used to increase the information transmitted in a communication context. In this paper, this idea is derived from a classical data transmission technique. Then, this model has been modified by controlling the transmitted power and by adapting the spectral coefficients of embedded codes according to the voice signal. This watermarking technique allows us to provide robust system to several treatment, specially to MP3 compression technique

1. INTRODUCTION

Le tatouage de signaux [1,7] a été dans un premier temps présenté comme une technique permettant de répondre au problème de la protection de droits d'auteur. L'idée de base consiste à ajouter une marque (encore appelée signature) aux données à protéger. La marque doit être imperceptible : dans le cas du tatouage audio la présence de la marque ne doit pas introduire de distorsion audible, et elle doit aussi être robuste vis-à-vis au traitements habituels du signal ainsi qu'aux attaques malveillantes. Cependant, cette marque doit être détectable par les personnes autorisées.

En plus de son utilisation pour la protection des droits d'auteur, les techniques de tatouage peuvent être envisagées pour la transmission de données [2,5] superposées à un signal audio. On parle alors de "contenu augmenté". Plusieurs applications sont possibles dans ce contexte. Par exemple, si l'information prioritaire à transmettre est de la parole, il

apparaît alors utile de transmettre des informations liées à une analyse vidéo tels que des paramètres caractérisant les lèvres en mouvement, le visage du locuteur ou toute autre information permettant d'augmenter l'intelligibilité de la conversation. Une telle application fait l'objet du projet RNRT ARTUS qui a pour objectif d'utiliser le tatouage de document pour insérer dans le flux audio une information inaudible issue de la vidéo (cf. Figure 1).

L'objectif de cet article est de présenter un schéma de tatouage d'audio applicable dans un contexte de communications numériques. Le cadre de ce travail nécessite des contraintes précises :

- Le débit de transmission du message inséré doit être le plus haut possible.
- Le système de tatouage doit également être robuste face à la compression de données, en particulier à la compression de type MPEG couche 3 (MP3) qui est souvent utilisée [4].
- Les attaques destinées à enlever la marque dans le signal de parole (piratage) ne seront pas prises en compte. Ceci car les signaux à transmettre ne sont pas des données sensibles.

2. IMPLEMENTATION

2.1. Cadre général

Pour aboutir à notre objectif, nous avons choisi d'utiliser une méthode de tatouage inspirée de la technique de communications CDMA (Code Division Multiple Access)[3]. Cette technique, qui permet la transmission de messages de plusieurs utilisateurs de

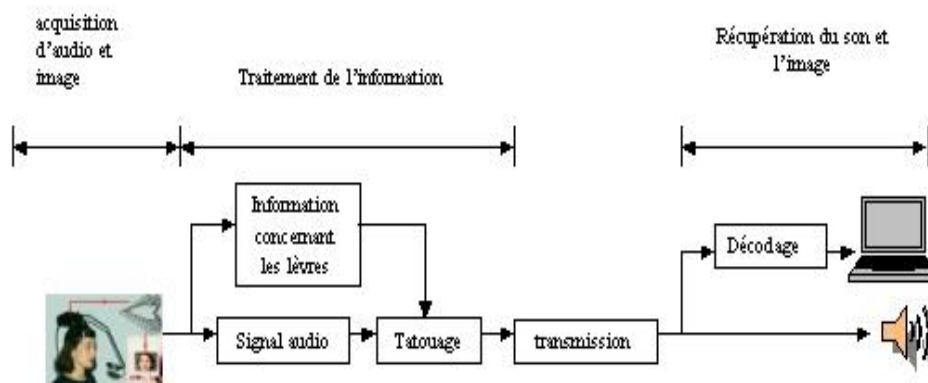


Figure 1. Rôle du tatouage au sien du Projet RNRT ARTUS

façon simultanée, peut être utilisée en systèmes mono-utilisateur pour les communications à faible rapport signal à bruit. La modulation utilisée consiste à remplacer chaque bit du message par un signal binaire déterminé, appelé code, caractérisé par ses propriétés de corrélation. Un code différent est assigné pour chaque utilisateur du système. Cette modulation, aussi appelée modulation par étalement de spectre, a la propriété de réaliser une distribution spectrale de puissance dans toute la largeur de bande disponible.

La démodulation s'effectue en calculant la corrélation entre le signal disponible en réception et une copie du code utilisé.

Pour notre implémentation, un seul type de code a été employé pour insérer un message binaire. La version positive du code correspondra au bit 1 du message et réciproquement la version négative sera utilisée pour le bit 0 (cf Figure 2).

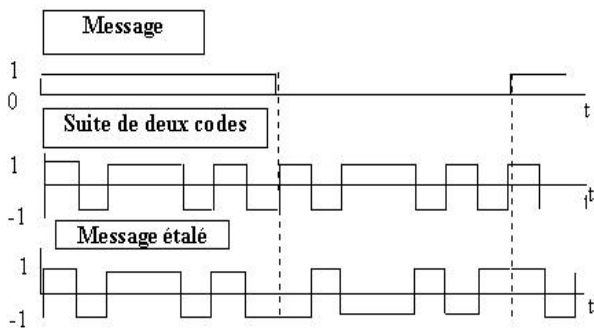


Figure 2 Modulation avec un seul code

Ceci correspond à la réalisation du calcul de :

$$(1) \quad W(t) = MSG(t) \times C(t).$$

où MSG(t) représente le message à envoyer, C(t) le code utilisé et W(t) le résultat de la modulation. Ce résultat est par la suite additionné au signal sonore :

$$(2) \quad S_w(t) = S(t) + W(t).$$

Dans ce type d'application il est important de préciser que la dynamique du signal ajouté doit être suffisamment faible pour que le code ne soit pas perceptible

A l'étape de réception, une copie du code est utilisée afin de décoder le signal reçu. Pour déterminer le message envoyé, il suffit de calculer le produit scalaire entre ce code et le message reçu. La décision sera prise en comparant sa valeur à zéro

La loi de décision est la suivante :

$$(3) \quad Si \langle S_w(t), C(t) \rangle > 0 \quad \Rightarrow \quad \text{Bit détecté} = 1$$

$$Si \langle S_w(t), C(t) \rangle < 0 \quad \Rightarrow \quad \text{Bit détecté} = 0$$

où $\langle S_w(t), C(t) \rangle$ correspond au produit scalaire entre les vecteurs $S_w(t)$ et $C(t)$. $\langle S_w(t), C(t) \rangle$ peut s'exprimer de la manière suivante:

$$(4) \quad \langle S_w(t), C(t) \rangle = \langle S(t), C(t) \rangle + \langle W(t), C(t) \rangle \\ \approx \langle W(t), C(t) \rangle$$

Dans ce contexte de communications W(t) correspond au message envoyé et S(t) (le son) est considéré comme du bruit.

Remarque : Pour réaliser le produit scalaire, il est nécessaire de connaître l'instant précis du premier bit du code. Cela est normalement réalisé dans une étape de synchronisation que nous ne décrivons pas.

2.2. Démarche adoptée

Après avoir défini la méthode de tatouage à mettre en oeuvre, des modifications ont été implantées afin d'améliorer les performances du schéma. Ces modifications comportent trois étapes (cf. Figure 3) qui sont détaillées plus amplement dans la suite du document:

1. *Le contrôle de la puissance du code.*
2. *La mise en forme fréquentielle des codes:* la densité spectrale des codes est modifiée de façon inverse à celle du signal sonore.
3. *Ajustement du contrôle de puissance:* en prenant en compte la mise en forme spectrale la puissance du code est recalculée par un système itératif

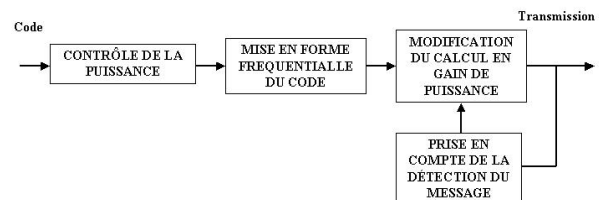


Figure 3 Démarche adoptée

2.3. Contrôle de la puissance

Dans un premier temps, nous avons décidé de contrôler la puissance de chacun des codes à ajouter. L'idée consiste à déterminer le coefficient minimal pour lequel chaque code doit être multiplié de telle manière à remplir les conditions (3). Nous avons décidé de multiplier le code W(t) par le gain K donné par:

$$(5) \quad K = - \frac{\langle S(t), C(t) \rangle}{\langle W(t), C(t) \rangle}$$

Le schéma de transmission est donc modifié de la manière suivante:

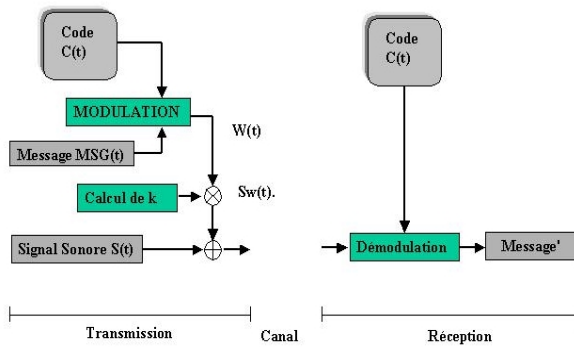


Figure 4 Schéma de transmission et tatouage

2.4. Mise en forme fréquentielle du code

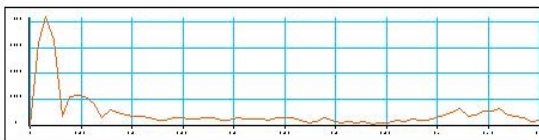
Le schéma présenté contient également une phase qui permet de diminuer au maximum l'effet du terme $\langle S(t), C(t) \rangle$ de l'équation (5).

Pour faire ceci, la stratégie adoptée a été de modifier la densité spectrale du code $C(t)$ en fonction de la densité spectrale de la parole $S(t)$. Nous avons choisi d'attribuer une distribution de la puissance du chaque code inverse à celle du signal audio. Le calcul des coefficients fréquentiels du code à insérer est réalisé grâce à une fonction logarithmique. Ceci a été utilisé pour permettre une pondération graduelle des coefficients. La phase des coefficients n'est pas modifiée. Le code résultant de cette opération est une fonction du $S(t)$ et $C(t)$. Le code modifié peut être exprimé de la façon suivante (cf. Figure 5) :

$$(6) \quad C_m(t) = f(S(t), C(t))$$

Nous aurons $W_m(t)$ au lieu de $W(t)$ à la sortie de l'étape de modulation. L'utilisation de cette technique permet de diminuer la puissance du message ajoutée au signal de parole.

Coefficients du Module d'un bloc S(f).



Coefficients du Module de Cm(f)

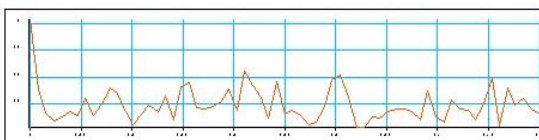


Figure 5 Distribution de puissance d'un code $C_m(f)$ de façon inverse à celle du signal audio de $S(f)$

2.5. Détection du message

Bien que nous disposions lors de la réception d'une copie du code original $C(t)$ pour détecter le message envoyé, l'allure des codes modifiés ($C_m(t)$) lors de l'insertion n'est pas disponible mais est pourtant nécessaire pour maximiser le résultat de la corrélation. Il devient donc important d'estimer $C_m(t)$ à la réception.

Cette estimation s'effectue à nouveau à partir des calculs de l'équation (6). Une différence est immédiatement remarquée: l'allure du code estimée à la réception n'est pas égale à celle du code utilisé à la transmission $C_m(t)$.

Ceci s'explique par le fait qu'à la réception le calcul (6) n'est pas fait à partir du $S(t)$ mais à partir du $S_w(t) = S(t) + W_m(t)$.

Cet à dire

$$C(t) = f(S(t), C(t)) \neq C_{mr}(t) = f(S_w(t), C(t))$$

Où $C_{mr}(t)$ est le code estimé en réception.

Afin de diminuer l'effet de cette différence, nous avons décidé de modifier le calcul du gain K (équation (5)). Ceci est expliqué dans la suite.

2.6. Ajustement de contrôle du gain de puissance

Le calcul du coefficient K (réalisé durant l'étape de transmission, équation (5)) a été donc modifié afin de diminuer l'effet produit par l'ajout du code. La démarche consiste alors à prédire dans l'étape de modulation le résultat de l'estimation en réception. Cette prédiction s'effectue en calculant de façon itérative le gain K_{i+1} à partir de $C_m(t)$ au lieu de $C(t)$. La formule permettant de calculer K_{i+1} est donc la suivante:

$$(7) \quad K_{i+1} = \frac{\langle S(t), C_{m,K_i}(t) \rangle}{\langle W_m(t), C_{m,K_i}(t) \rangle} \text{ avec } K_0 = K$$

Dans la pratique un nombre d'itération égale à 3 permet d'obtenir une valeur stable.

3. RESULTATS

Nous avons appliqué cette méthode à des signaux de parole et nous avons évalué le résultat obtenu vis-à-vis du débit de transmission du message. Les signaux ont été échantillonnés à 16 kHz sur 16 bits et chacun des fichiers résultants à une durée de dix secondes. Notre implémentation permet de garantir une détection exempte d'erreurs dans le cas d'un canal idéal sans bruit additif.

En ce qui concerne l'audibilité du message, deux personnes ont écouté les fichiers marqués et non marqués et ils ont évalué de façon subjective la perceptibilité de la marque. Ce test a montré qu'il reste imperceptible jusqu'à 125 bits/s. A partir de cette fréquence une distorsion faible commence à être perçue.

Nous avons testé aussi la robustesse vis-à-vis la compression MPEG MP3 pour un débit de 128 kbits/s. La détection de du message s'effectue ensuite sur les fichiers décompressés. Nous avons obtenu de taux d'erreur faibles dans les fichiers traités. Le tableau 1 montre le nombre d'erreurs obtenu. Nous avons traité douze fichiers. Deux d'entre eux ont suivi les tests de compression.

Débit bits/s	31,25	62,5	125	250	500
Sans Compression	0	0	0	0	0,06
MP3	0,00	0,00	0,55	1,51	3,16

Tableau 1: Proportion de bits erronés (en %) avec et sans compression.

4. CONCLUSIONS ET PERSPECTIVES

Nous avons proposé d'augmenter le contenu d'un signal de parole lors d'une transmission numérique à partir d'un système de tatouage fondé sur la technique d'étalement de spectre. Le schéma résultant a été amélioré en tenant compte des conditions imposées par le signal audio. Nous avons implémenté des systèmes de contrôle de puissance du code. L'adaptation à l'allure fréquentielle des sons a également été prise en compte. Le système présente des résultats acceptables face à la compression MP3, l'étude du modèle auditif humain et la méthode de quantification utilisée dans le compresseur [6] permettront d'améliorer le schéma.

Dans ce travail, nous avons focalisé nos efforts sur la quantité d'information qui peut être additionnée avec ce schéma sans prendre en compte le problème de perte de synchronisation. Des travaux, basés sur l'utilisation de symboles synchronisant, sont en cours et permettront d'approcher au plus près les conditions d'une transmission numérique réelle.

BIBLIOGRAPHIE

- [1] P. Bas. Méthodes de tatouage d'images fondées sur le contenu. Thèse LIS-INPG, Grenoble, septembre 2000.
- [2] Transmission d'un message numérique caché dans un signal audio. Joël Lienard. Colloque Gretsi 2001, Toulouse, France.
- [3] Proakis. Digital Communications. Mc. Graw Hill. 1995.
- [4] Site internet du Fraunhofer Institut. <http://www.iis.fhg.de/amm/techinf/layer3/index.html>
- [5] L. Boney, A. H. Tewfik and K. N. Hamdy. Digital Watermarks for Audio. In Eusipco, pages 473-480, Trieste, 1996.
- [6] A review of algorithms for perceptual coding of digital audio signals. Painter, T. Spanias, A. Arizona State University.
- [7] C. Neubauer and J. Herre, "Audio watermarking of MPEG-2 AAC bit streams" In 108th Convention AES, Paris, February 2000.