Tatouage audio exploitant des propriétés de cyclostationnarité

Cyclostationarity-based audio watermarking

par Leandro de C. T. GOMES[1], Mamadou MBOUP[1], Madeleine BONNET[1], et Nicolas MOREAU[2]
[1]UFR de mathématiques et informatique - CRIP5/InfoCom Université René Descartes (Paris V)
45, rue des Saints-Pères, 75270 Paris Cedex 06 FRANCE
Tel.: +33 1 44 55 35 24 Fax: +33 1 44 55 35 35 E-mail: {tgomes, mboup, bonnet}@math-info.univ-paris5.fr
[2]ENST/TSI
46, rue Barrault, 75634 Paris Cedex 13 FRANCE
Tel.: +33 1 45 81 77 33 Fax: +33 1 45 88 79 35 E-mail: moreau@tsi.enst.fr

résumé et mots clés

Le tatouage audio consiste à insérer une information inaudible dans un signal. Cette information est généralement représentée par un signal pseudo-aléatoire, le tatouage, détecté à l'aide d'une mesure de corrélation. Si le tatouage doit être robuste à des attaques malveillantes, le signal pseudo-aléatoire est impérativement secret et la détection est dite privée. Nous présentons une approche qui utilise comme tatouage un signal cyclostationnaire. Tout en étant détectable de façon privée par corrélation, le tatouage peut aussi être détecté publiquement grâce à la propriété de cyclostationnarité. Le choix judicieux de suites cyclostationnaires permet de cacher à la fois des données privées et publiques dans le signal.

tatouage audio, protection de droits d'auteur, tatouage public, cyclostationnarité.

abstract and key words

Audio watermarking consists in embedding inaudible information in a signal. This information is generally represented by a pseudorandom signal, the watermark, detected by means of a correlation measure. If robustness to malicious attacks is required, the pseudorandom signal must be secret and the detection is private. We present an approach that uses a cyclostationary signal as the watermark. While still privately detectable through correlation, the watermark may also be publicly detected by exploiting its property of cyclostationarity. The suitable choice of cyclostationary sequences provides for hiding both private and public data in the signal.

audio watermarking, copyright protection, public watermarking, cyclostationarity.

1. introduction

1.1. tatouage : définition et applications

Les signaux audio sous forme numérique étant très facilement reproductibles sans aucune distorsion, des techniques de protection efficaces sont devenues indispensables. Cette nécessité a été renforcée par les nouveaux moyens de distribution, comme Internet, ainsi que par les méthodes de compression, telle la norme MPEG-2 AAC. Le tatouage a été proposé pour résoudre ce problème.

Le tatouage est un signal que l'on insère dans le signal audio. Le signal tatoué contient alors des informations (données cachées) qui peuvent être utilisées à plusieurs fins. Par exemple, elles peuvent servir à identifier le propriétaire (protection du copyright), à identifier la source de copies illicites, à vérifier l'intégrité d'un document, à suivre la trace d'un signal dans un réseau, à insérer des informations complémentaires telles que le titre, l'auteur, le numéro ISBN.

Les contraintes que doit satisfaire un système de tatouage audio dépendent de l'application. Les principales contraintes sont :

• l'inaudibilité : le signal de tatouage ne doit pas être perçu par l'auditeur ;

• la robustesse : le tatouage doit résister à toute modification du signal, tant que cette modification n'entraîne pas de dégradation de la qualité ;

• la fiabilité : le système doit présenter un taux élevé de bonne détection ainsi qu'un faible taux de fausse alarme.

Le tatouage doit résister à des opérations licites comme la compression/décompression, le filtrage et le ré-échantillonnage. Pour la protection du copyright, la résistance à des attaques intentionnelles doit être assurée. Cependant, pour des applications d'intégrité, le tatouage doit être fragile.

La cryptographie, qui permet de sécuriser un document, présente une différence majeure avec le tatouage. Après décryptage, le document n'est plus sécurisé alors, que la protection apportée par le tatouage peut persister indéfiniment.

1.2. schéma de tatouage

La figure 1 illustre un schéma générique de tatouage. Une clé (clé 1), connue du propriétaire seul (clé privée), est utilisée pour générer le tatouage. Une autre clé (clé 2) est utilisée pour la détection. Quand ces deux clés sont identiques, le schéma de tatouage est dit symétrique, sinon il est dit dissymétrique. Si la clé 2 est connue publiquement (clé publique), elle doit permettre à l'utilisateur de vérifier la présence du tatouage mais non pas d'isoler le tatouage du signal ; le schéma de tatouage est alors dit public. Si la clé 2 est aussi secrète, le schéma est dit privé. Un schéma symétrique est forcément privé, tandis qu'un schéma public est forcément dissymétrique.

Dans la phase de détection, l'utilisateur vérifie d'abord si le signal est ou non tatoué (détection binaire). Si le tatouage est présent, l'utilisateur essaie d'extraire l'information contenue dans le tatouage (détection de données cachées). Les détections binaire et des données cachées peuvent être privées ou publiques.

1.3. état de l'art

À notre connaissance, le premier papier sur le tatouage audio a été publié en 1996 [1,2]. Une séquence pseudo-aléatoire (la clé privée) est ajoutée au signal audio. Afin de garantir l'inaudibilité du tatouage, la séquence est mise en forme spectralement suivant un seuil de masquage obtenu à l'aide d'un modèle psychoacoustique [3,4]. La détection est réalisée par une mesure de corrélation entre le tatouage original et un tatouage estimé à partir du signal observé. Cette détection est évidemment privée.

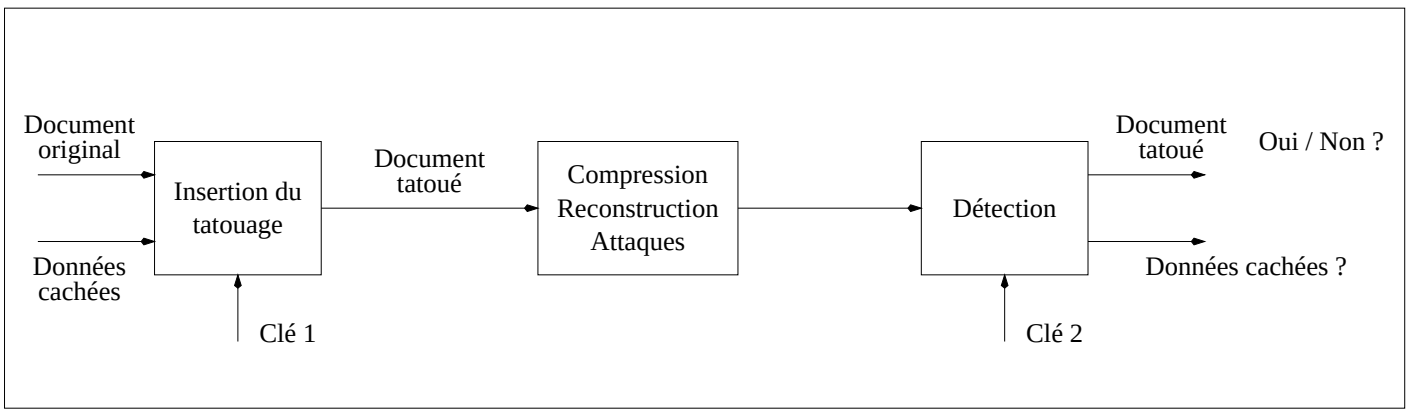


Figure 1. – Schéma générique de tatouage

Une méthode à détection publique a été proposée récemment dans [5,6]. Le signal de tatouage présente une densité spectrale de puissance publiquement connue, la détection consiste alors à rechercher cette propriété dans le signal tatoué. Remarquons que la seule connaissance de la densité spectrale de puissance ne suffit pas pour reconstruire le signal de tatouage.

D'autres techniques ont été présentées dans la littérature. Par exemple dans [7], le tatouage est inséré directement dans un train binaire MPEG-AAC par modification des facteurs d'échelle. Dans [8], les auteurs proposent une technique similaire à celle présentée dans [1] mais en opérant aussi sur le train binaire. Dans [9], bien que la méthode présentée soit une méthode cryptographique, l'utilisateur n'a jamais accès à la version originale non protégée du document.

Dans [10], la méthode de tatouage est fondée sur une modulation BPSK et sur un étalement en fréquence. Les composantes spectrales du signal audio en dessous du seuil de masquage sont remplacées par des composantes de tatouage. Dans [11], ce sont les propriétés de masquage temporel qui sont exploitées pour rendre le tatouage inaudible : on utilise une sorte d'écho du signal original en tant que signal de tatouage.

1.4. plan du papier

Nous nous intéressons ici au problème du tatouage public. Après avoir vu le tatouage comme étant une chaîne de communication, nous présentons notre schéma de tatouage privé. Nous passons ensuite au tatouage public. Pour cela, nous introduisons dans le signal tatoué une propriété, la cyclostationnarité, qui peut être détectée à l'aide d'une clé publique. Nous montrons alors comment combiner tatouages privé et public. Des résultats expérimentaux illustrent la méthode proposée.

2. tatouage vu comme une chaîne de communication

Le processus de tatouage peut être vu comme une chaîne de communication bruitée [10] : le tatouage est le signal à transmettre et le signal audio (ainsi que d'éventuelles distorsions subies par le signal tatoué) est considéré comme un bruit beaucoup plus puissant que le tatouage (à cause de la condition d'inaudibilité). Cette approche est illustrée figure 2.

L'information contenue dans le tatouage est représentée par des symboles. On construit un dictionnaire $C = [\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{L-1}]$ contenant L vecteurs de longueur N, chacun associé à un symbole. Le modulateur reçoit une suite de symboles et produit un signal $s(n)$ par concaténation des vecteurs correspondant à ces symboles. Le débit-symbole R_s est égal à la fréquence d'échantillonnage divisée par la longueur des vecteurs du dictionnaire. Le débit binaire est égal au débit-symbole multiplié par le nombre de bits par symbole.

Le filtre $H(f)$, issu d'un modèle psychoacoustique (noté « MPA » sur la figure 2), met en forme spectralement le signal de tatouage. Cette opération permet d'avoir un rapport signal à tatouage de l'ordre de 20 dB sans distorsion audible. Ce rapport est relativement constant au cours du temps et quelque soit le type de musique. Le signal résultant $w(n)$ est alors additionné au signal $x(n)$, produisant le signal tatoué $y(n)$.

Les distorsions que subit le tatouage sont en général colorées, c'est le cas du signal audio lui-même et du bruit introduit par une compression/décompression.

Le signal tatoué observé $\hat{y}(n)$ est d'abord filtré par $G(f)$, filtre de Wiener servant à augmenter le rapport de puissance entre le tatouage et le signal audio original. Le filtre est estimé à partir

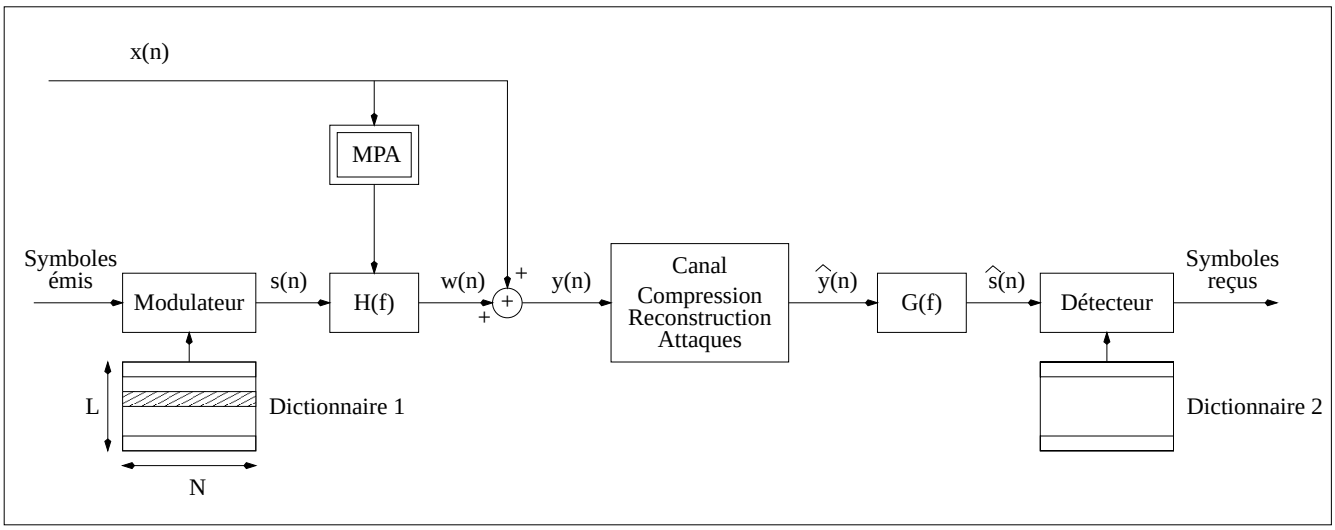


Figure 2. – Tatouage vu comme une chaîne de communication

de $\hat{y}(n)$, puisque le signal original $x(n)$ n'est pas disponible à la détection. La sortie $\hat{s}(n)$ est une estimation de $s(n)$. Le détecteur reçoit $\hat{s}(n)$ et, à l'aide d'un second dictionnaire, il essaie d'identifier les symboles transmis.

3. tatouage privé

Le schéma de la figure 2 est directement utilisé pour le tatouage privé. Les deux dictionnaires sont identiques (le schéma de tatouage est symétrique) et correspondent à la clé privée.

3.1. tatouage à détection binaire

Bien que fondé sur la même idée que [1], le schéma de tatouage à détection binaire privé que nous présentons ne requiert pas le signal audio $x(n)$ pour la détection.

Le dictionnaire $C = [\mathbf{v}]$ ne contient qu'un vecteur transmis continûment[1]. La détection est effectuée sur des fenêtres successives du signal $\hat{y}(n)$ de longueur égale à celle du vecteur du dictionnaire. Pour chaque fenêtre, le test d'hypothèse est le suivant :

• $\mathbf{H_0}$: $\hat{s}(n) = b(n)$

• $\mathbf{H_1}$: $\hat{s}(n) = ks(n) + b(n)$

où k est une constante et $b(n)$ représente toutes les distorsions subies par le tatouage, incluant le signal audio lui- même. Si le signal n'est pas tatoué (hypothèse $\mathbf{H_0}$), $\hat{s}(n)$ est égal à $b(n)$; sinon (hypothèse $\mathbf{H_1}$), $\hat{s}(n)$ contient une composante proportionnelle à $s(n)$. La décision concernant la présence du tatouage est prise à l'aide d'une mesure de corrélation comparée à un seuil.

La fiabilité de la détection dépend fortement du choix du seuil, qui est lié à l'application. Si une faible probabilité de fausse alarme est requise, le seuil doit être élevé (ceci réduit alors la probabilité de bonne détection), tandis qu'une forte probabilité de bonne détection est obtenue avec un seuil plus faible (ceci augmente la probabilité de fausse alarme). Un compromis est donc recherché.

3.2. données cachées

L'utilisation d'un dictionnaire à plusieurs vecteurs $C = [\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{L-1}]$ permet au tatouage de transporter plus de bits d'information. La détection est encore réalisée pour des fenêtres de longueur égale à celle des vecteurs du dictionnaire[2].

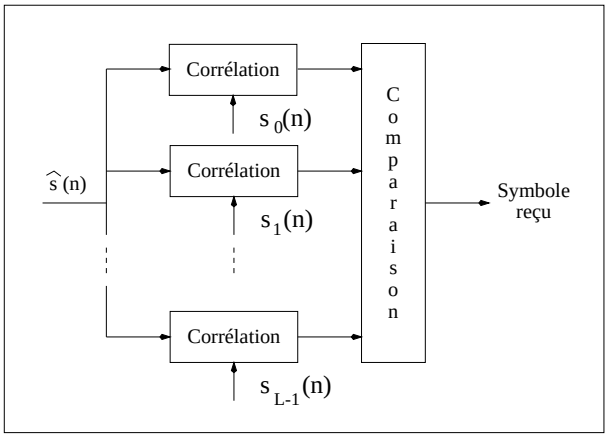


Figure 3. – Détection privée pour données cachées

Pour chaque fenêtre du signal observé, le test d'hypothèse s'énonce :

• $\mathbf{H_0}$: $\hat{s}(n) = ks_0(n) + b(n)$

• $\mathbf{H_1}$: $\hat{s}(n) = ks_1(n) + b(n)$

$\vdots$

• $\mathbf{H_{L-1}}$: $\hat{s}(n) = ks_{L-1}(n) + b(n)$

où $s_i(n)$ est un signal à durée limitée dont les échantillons sont identiques à ceux du vecteur v_i du dictionnaire.

Pour la détection, L mesures de corrélation sont calculées selon le schéma de la figure 3.

Tous les vecteurs doivent avoir la même norme (*i.e.* la même énergie), afin que les mesures de corrélation puissent être directement comparées. Le symbole reçu est celui associé à la plus forte corrélation.

4. tatouage public

Nous introduisons maintenant l'utilisation de la cyclostationnarité pour le tatouage à détection publique.

4.1. cyclostationnarité

Soit un processus aléatoire $X(t)$, non-stationnaire, à temps continu, dont le moment d'ordre k est $m_k(t)$. Ce processus est dit *cyclostationnaire* [14] de cyclo-fréquence α si ses moments satisfont

$$m_k(t + 1/\alpha) = m_k(t)$$

quelque soit t. Si cette propriété est vraie pour $k \leqslant 2$, le processus est dit cyclostationnaire au sens large. Un processus cyclostationnaire discret peut être obtenu par échantillonnage de $X(t)$. En télécommunications, des processus générés par

1. En principe, un dictionnaire à un seul vecteur ne peut transmettre de l'information ; cependant, l'idée du tatouage à détection binaire est de vérifier la présence ou non d'un tatouage, ce qui correspond à un bit d'information.

2. Les problèmes de synchronisation entre émetteur et récepteur sont traités dans [12] et [13].

modulation d'amplitude, de fréquence ou de phase sont cyclostationnaires de cyclo-fréquence égale à la fréquence porteuse.
Un signal est dit cyclostationnaire s'il est une réalisation d'un processus cyclostationnaire. Les signaux cyclostationnaires présentent des moyennes temporelles périodiques de fréquence égale à la cyclo-fréquence[3].

4.2. détection de la cyclostationnarité

Les signaux cyclostationnaires peuvent être caractérisés par la *propriété de génération de lignes spectrales* [14,15]. Si l'on applique une transformation quadratique à un signal cyclostationnaire (au sens large) de cyclo-fréquence α, le signal résultant contient une composante sinusoïdale non-infinitésimale de fréquence 2α. (L'inverse étant vrai : si un signal, obtenu par-transformation quadratique, présente une composante sinusoïdalenon-infinitésimale de fréquence 2α, le signal original est cyclostationnaire de cyclo-fréquence α). Un signal cyclostationnaire $x(n)$ peut être obtenu à partir d'un signal stationnaire centré $a(n)$ en le multipliant par une sinusoïde :

$$x(n) = a(n)\cos(2\pi\alpha n).$$

En élevant $x(n)$ au carré, on a :

$$y(n) = x^2(n) = a^2(n)\cos^2(2\pi\alpha n)$$
$$= \frac{1}{2}a^2(n) + \frac{1}{2}a^2(n)\cos(4\pi\alpha n).$$

Une composante de fréquence 2α apparaît dans $y(n)$. Cette propriété peut être exploitée pour détecter la cyclostationnarité. Plutôt qu'utiliser un simple carré comme dans l'exemple précédent, une transformation quadratique optimisée peut être obtenue par une représentation en série de Volterra du second ordre. Soit $x(n)$ un signal nul en dehors de l'intervalle $[0, M-1]$ et

$$y_k(n) = x(n)x(n-\tau_k), \quad k = 0, 1, \ldots, L-1$$

où $\tau_k = k + \dfrac{1-L}{2}$ et $1 \leqslant L \leqslant 2M-1$. La série de Volterra représente la transformation du second ordre comme une combinaison linéaire de tels produits dont les coefficients sont calculés par estimation des moindres carrés. Le but étant de rendre le résultat de la transformation du second ordre aussi proche que possible d'une exponentielle complexe de fréquence 2α, nous définissons

$$\mathbf{y}(n) = [y_0(n), y_1(n), \ldots, y_{L-1}(n)]^T$$

et

$$\mathbf{w} = [w_0, w_1, \ldots, w_{L-1}]^T.$$

En annulant le gradient du critère de minimisation :

$$\widetilde{\mathbf{w}} = \arg \min_{\mathbf{w}} \sum_{n=0}^{M-1} \mid \mathbf{w}^T\mathbf{y}(n) - \exp(j4\pi\alpha n) \mid^2$$

nous obtenons :

$$\sum_{i=0}^{L-1}[\widetilde{w}_i \sum_{n=0}^{M-1} y_k(n)y_i(n)] = \sum_{n=0}^{M-1} y_k(n)\exp(j4\pi\alpha n),$$
$$k = 0, 1, \ldots, L-1.$$

En définissant

$$R_{ki} = \sum_{n=0}^{M-1} y_k(n)y_i(n)$$

et

$$p_k = \sum_{n=0}^{M-1} y_k(n)\exp(j4\pi\alpha n)$$

où R_{ki} est l'élément de la ligne k et de la colonne i de la matrice $\mathbf{R}$, et p_k le $k^{\text{ème}}$ élément du vecteur colonne $\mathbf{p}$, nous avons :

$$\mathbf{R}\widetilde{\mathbf{w}} = \mathbf{p}$$

et

$$\widetilde{\mathbf{w}} = \mathbf{R}^{-1}\mathbf{p}.$$

La matrice $\mathbf{R}$ est supposée non singulière. Le résultat de la transformation optimale du second ordre donne

$$u(n) = \widetilde{\mathbf{w}}^T\mathbf{y}(n).$$

Un coefficient de détection peut être obtenu par corrélation de $u(n)$ et d'une exponentielle complexe de fréquence 2α :

$$Y = \frac{1}{M}\sum_{n=0}^{M-1} u(n)\exp(j4\pi\alpha n).$$

La valeur du coefficient Y est comprise entre 0 ($u(n)$ n'a pas de composante spectrale de fréquence 2α) et 1 (une exponentielle complexe de fréquence 2α est parfaitement reconstruite). En pratique, dû à la faible puissance du tatouage par rapport à celle du signal audio, ce coefficient se situe généralement dans l'intervalle allant de 0 à 0,1. Si Y est au-dessus d'un seuil donné, on considère que $x(n)$ contient une composante cyclostationnaire de cyclo-fréquence α. Ce seuil, déterminé expérimentalement dans notre système, est fortement dépendant du signal audio considéré, ce qui pose un problème de fiabilité de la détection binaire. Ce problème ne se pose pas lorsque l'on présume que le signal contient une composante cyclostationnaire et le but est de vérifier si cette composante a une cyclo-fréquence α ou β. Dans ce cas, la détection est réalisée pour chacune des cyclo-fréquences et les coefficients obtenus sont simplement comparés.

4.3. tatouage à détection binaire

Le tatouage détectable de façon publique consiste en un signal cyclostationnaire incrusté au signal audio. Seule la cyclo-fréquence (la clé publique) est connue. La technique décrite précé-

3. Un signal issu d'un processus stationnaire peut présenter des moyennes temporelles périodiques. Tel est le cas lorsqu'un processus cyclostationnaire est rendu stationnaire par l'introduction d'une phase aléatoire. Dans ce cas, on dit que le signal présente de la *cyclostationnarité cachée* [14].

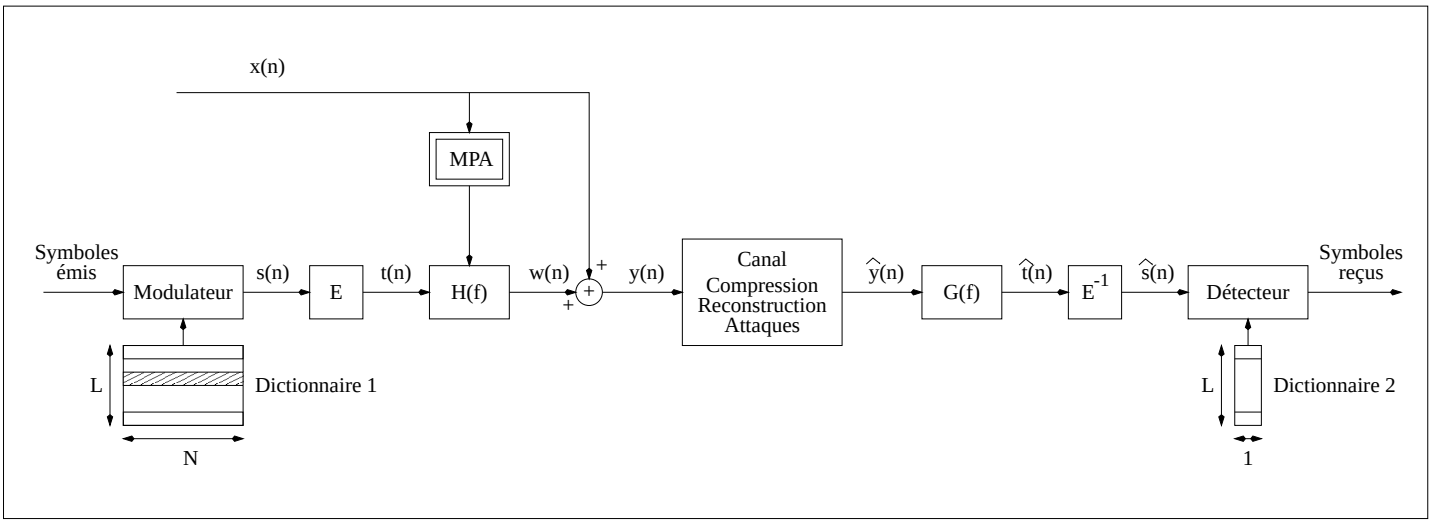


Figure 4. – Schéma de tatouage public.

demment permet à l'utilisateur de vérifier si un signal cyclostationnaire est présent ou non sans pour autant permettre de reconstruire ce signal (qui est la clé privée)[4]. L'utilisateur est donc en mesure de vérifier si le signal est tatoué ou non, mais il n'est pas capable de récupérer (et supprimer) le tatouage.

La figure 4 illustre notre schéma de tatouage public. L'idée du tatouage vu comme une chaîne de communication (section 2) est à nouveau utilisée, mais les dictionnaires 1 et 2 sont différents l'un de l'autre. Pour le tatouage à détection binaire, le dictionnaire 1 contient un vecteur qui est transmis continûment. Les éléments de ce vecteur correspondent aux échantillons d'un signal cyclostationnaire de cyclo-fréquence α. Le dictionnaire 2, à son tour, ne contient que les valeurs des cyclo-fréquences. Aucune autre information n'est fournie au détecteur.

Le signal est d'abord étalé en fréquence[5] à l'aide d'un entrelaceur (« E » sur la figure) qui permute aléatoirement les échantillons de chaque fenêtre. Afin de garantir l'inaudibilité du tatouage, le signal est alors mis en forme spectralement à l'aide d'un seuil de masquage issu d'un modèle psychoacoustique. Le signal tatoué $y(n)$ résultant peut subir des opérations de compression/décompression, filtrage etc. Il peut aussi être soumis à des attaques essayant d'effacer le tatouage.

Dans la phase de détection, le signal tatoué observé entre d'abord dans le filtre de Wiener. Il subit alors l'entrelacement inverse de celui de l'émetteur. Remarquons que cette dernière opération a aussi pour effet de blanchir le signal audio.

Comme pour le tatouage privé, le problème est formulé comme un test d'hypothèse appliqué à chaque fenêtre du signal observé :

• $\mathbf{H_0}$: $\hat{s}(n) = b(n)$

• $\mathbf{H_1}$: $\hat{s}(n) = b(n) +$ composante cyclostationnaire de cyclo-fréquence α

où $\hat{s}(n)$ est l'estimation de la sortie du modulateur et $b(n)$ représente toutes les distorsions subies par le tatouage, incluant le signal audio lui-même. Si le signal n'est pas tatoué (hypothèse $\mathbf{H_0}$), $\hat{s}(n)$ est égal à $b(n)$; sinon (hypothèse $\mathbf{H_1}$), $\hat{s}(n)$ contient une composante cyclostationnaire de cyclo-fréquence α.

L'étalement en fréquence est nécessaire car le signal cyclostationnaire est à bande étroite. Bien qu'un signal à large bande puisse être utilisé, la détection est facilitée pour un signal à bande étroite (tant que sa puissance totale est conservée). Ceci résulte du fait qu'à l'intérieur de la bande de fréquence occupée par le signal cyclostationnaire, le rapport de puissance (signal cyclostationnaire / signal audio) augmente lorsque la largeur de bande décroît. Les composantes spectrales en dehors de cette bande de fréquence peuvent être éliminées facilement par filtrage, bien que leur influence sur la détection soit négligeable. La largeur de bande du signal cyclostationnaire ne peut cependant pas être trop réduite, puisqu'il serait facile pour les pirates de détruire un tatouage proche d'une sinusoïde. Autrement dit, la clé publique (la cyclo-fréquence) contiendrait presque toute l'information sur la clé privée (le signal cyclostationnaire).

De fausses alarmes peuvent être provoquées par une éventuelle caractéristique cyclostationnaire du signal audio. Ces cas sont rares, puisque le détecteur recherche la présence d'une cyclo-fréquence très particulière. Néanmoins, l'opération inverse d'étalement du spectre devrait détruire toute caractéristique cyclostationnaire du signal audio.

4. Bien que des techniques de filtrage non linéaire puissent être utilisées pour essayer d'isoler le signal cyclostationnaire, ceci n'est pas possible en général car l'utilisateur ne connaît pas les propriétés statistiques de ce signal (seule la cyclo-fréquence est connue).

5. Cette opération serait inutile si le signal cyclostationnaire était blanc ; comme nous l'expliquerons plus tard, les performances de la détection sont d'autant améliorées que la largeur de bande du signal cyclostationnaire est étroite.

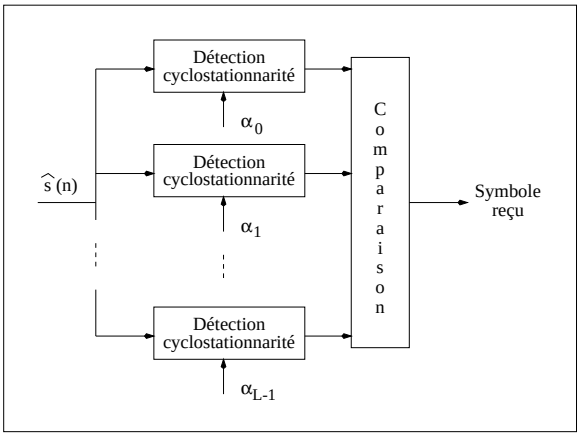


Figure 5. – Détection publique des données.

4.4. données cachées

Le schéma précédent peut être étendu pour insérer des données à destination publique dans le signal audio. On construit un dictionnaire $C = [\mathbf{v}_0, \mathbf{v}_1, \cdots, \mathbf{v}_{L-1}]$ dont les éléments des L vecteurs correspondent aux échantillons de L signaux cyclostationnaires $s_0(n), s_1(n), \ldots, s_{L-1}(n)$ de cyclo-fréquences $\alpha_0, \alpha_1, \ldots, \alpha_{L-1}$. Le signal $s(n)$ est généré par concaténation des vecteurs correspondant aux symboles transmis. Le signal cyclostationnaire subit les mêmes opérations que pour la détection binaire avant d'être ajouté au signal audio. La détection est maintenant réalisée pour chacune des L cyclo-fréquences possibles. On obtient donc L coefficients de détection et on choisit comme symbole reçu celui associé au plus fort de ces coefficients. La comparaison directe entre les coefficients est rendue possible car tous les signaux du dictionnaire ont la même puissance. La figure 5 représente un tel détecteur.

5. combinaison d'un tatouage public et d'un tatouage privé

Nous allons montrer comment des informations à caractère public et d'autres à caractère privé peuvent coexister dans un même tatouage.

5.1. construction du dictionnaire

Le dictionnaire qui va contenir à la fois les symboles publics et les symboles privés (dictionnaire mixte) est représenté par une matrice où les L_{pb} colonnes correspondent aux symboles publics et les L_{pv} lignes aux symboles privés :

$$C = \begin{bmatrix} \mathbf{v}_0^0 & \mathbf{v}_0^1 & \cdots & \mathbf{v}_0^{L_{pb}-1} \\ \mathbf{v}_1^0 & \mathbf{v}_1^1 & \cdots & \mathbf{v}_1^{L_{pb}-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_{L_{pv}-1}^0 & \mathbf{v}_{L_{pv}-1}^1 & \cdots & \mathbf{v}_{L_{pv}-1}^{L_{pb}-1} \end{bmatrix}.$$

Ce dictionnaire comporte des signaux cyclostationnaires de même énergie. Les signaux situés dans une même colonne ont la même cyclo-fréquence et correspondent donc à un même symbole public. Cependant, ces signaux sont générés à partir de différents signaux stationnaires (grâce à une modulation d'amplitude), correspondant ainsi à différents symboles privés. En revanche, tous les signaux d'une même ligne sont issus d'un même signal stationnaire mais présentent des cyclo-fréquences différentes ; ils correspondent donc au même symbole privé mais à différents symboles publics.

5.2. détection du tatouage

La détection des symboles publics est effectuée de la même façon que celle décrite en 4.4. La clé publique est l'ensemble des cyclo-fréquences, ce qui permet à l'utilisateur d'extraire les données publiques incluses dans le tatouage. La détection des symboles privés est obtenue par corrélation, la clé privée étant le dictionnaire mixte. Chaque fenêtre du signal observé est corrélée avec les $L_{pv} \times L_{pb}$ entrées du dictionnaire et l'on choisit le symbole correspondant à la plus forte corrélation. Les données privées, aussi bien que les données publiques, peuvent être trouvées par cette procédure puisque que l'on peut identifier la ligne et la colonne correspondant au symbole du dictionnaire.

6. résultats expérimentaux

6.1. conditions d'expériences

Pour générer un signal cyclostationnaire $s(n)$ de cyclo-fréquence α, un signal stationnaire, centré, blanc, gaussien $a(n)$ passe dans un filtre passe-bas de fréquence de coupure 500 Hz. Ce signal est ensuite modulé en amplitude par une porteuse de fréquence α :

$$s(n) = a(n)\cos(2\pi\alpha n).$$

On utilise des fenêtres de 512 échantillons sans recouvrement, ce qui correspond à la longueur des vecteurs du dictionnaire. Chaque vecteur est répété 5 fois, ce qui donne une longueur totale de 2 560 échantillons par symbole transmis. La fréquence

d'échantillonnage est de 32 kHz. Le dictionnaire a pour dimensions $L_{pv} = 128$ et $L_{pb} = 16$, c'est-à-dire 128 suites gaussiennes presque orthogonales et 16 cyclo-fréquences (uniformément distribuées entre 2 kHz et 6 kHz). Ainsi, chaque vecteur du dictionnaire contient 7 bits d'information privée et 4 bits d'information publique, ce qui donne un débit binaire de 87,5 bit/s pour les données privées et 50 bit/s pour les données publiques. Afin de simplifier les simulations, il n'y a pas eu de recouvrement entre fenêtres ; pour des applications pratiques, l'utilisation de fenêtres avec recouvrement pourrait apporter une amélioration de la qualité du signal perçu.

Comme dans la figure 4, on utilise un entrelaceur pour blanchir le signal cyclostationnaire. Les échantillons sont permutés de façon aléatoire pour chaque fenêtre de 512 échantillons. Un filtre de mise en forme spectrale est utilisé pour ajuster la densité spectrale de puissance du signal entrelacé au seuil de masquage fourni par le modèle psychoacoustique MPEG-2 numéro 1. La puissance du tatouage est en moyenne 20 dB en dessous de celle du signal audio.

Le filtre de Wiener $G(f)$ ainsi que le seuil de masquage sont estimés à partir du signal observé $\hat{y}(n)$. Cette estimation n'altère pas trop le seuil de masquage puisque le tatouage $w(n)$ est beaucoup plus faible que le signal audio original $x(n)$. Après entrelacement inverse, les 5 répétitions du signal cyclostationnaire sont moyennées afin d'augmenter le rapport tatouage à signal audio.

6.2. détection binaire

La figure 6 présente les résultats de détection publique pour le signal test « svega » (« Tom's diner » (version a cappella) de Suzanne Vega, 9,4 s). Pour une probabilité de fausse alarme de 0,01, la probabilité de décision correcte est proche de 0,9. Des résultats similaires ont été obtenus pour deux autres signaux tests (« violin » , concerto de Bach pour violon, 26,2 s, et « baron », un morceau de musique des Caraïbes de Baron, 11,5 s). Pour un quatrième signal test (« piano », un morceau de piano de jazz par Petrucciani, 13,6 s), un plus mauvais résultat a été obtenu ; nous recherchons actuellement la cause de cette mauvaise performance.

Pour la détection privée binaire, le tatouage a été correctement détecté pour toutes les fenêtres des signaux tests.

6.3. données cachées

Le tableau 1 montre les taux d'erreur de détection pour les données cachées. Aucune erreur n'est apparue pour la détection des données privées ou publiques en utilisant la clé privée. En utilisant la clé publique, les taux d'erreur trouvés sont assez faibles (entre 0,03 et 0,09). De même que pour la détection binaire, le signal « piano » n'a pas donné de bons résultats.

7. conclusion

Nous avons présenté une technique de tatouage fondée sur des propriétés de cyclostationnarité. Notre technique permet d'insérer dans un signal audio à la fois des données publiques et des données privées. Les données publiques, accessibles à tous, sont

Tableau 1. – Taux d'erreur pour détection privée et publique

Signal	Durée	Taux d'erreur : Données publiques, Détection publique	Taux d'erreur : Données privées, Détection privée	Taux d'erreur : Données publiques, Détection privée
svega	9,4 s	0,03	0,00	0,00
violin	26,2 s	0,09	0,00	0,00
baron	11,5 s	0,03	0,00	0,00

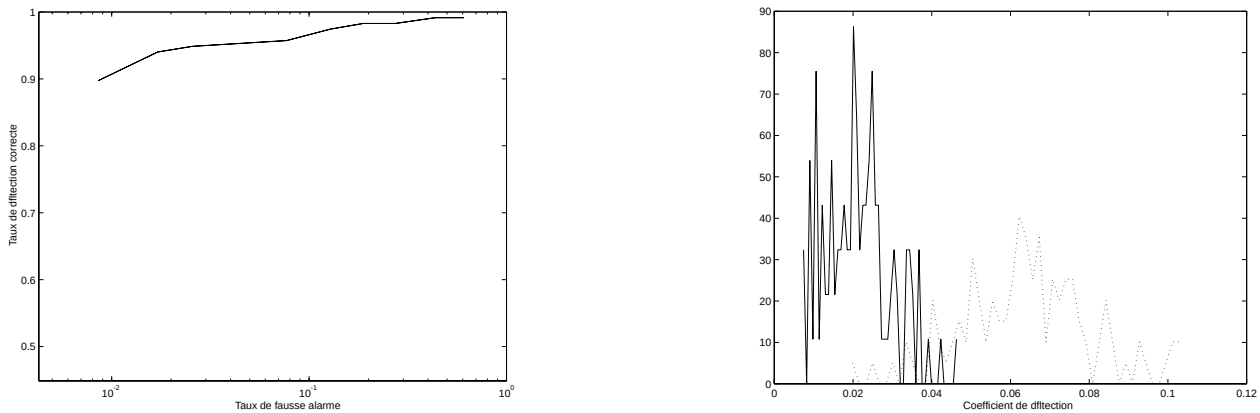


Figure 6. – Résultats de détection binaire pour le signal « svega ». À gauche : probabilité de détection correcte en fonction de la probabilité de fausse alarme. À droite : histogramme du coefficient de détection sur des fenêtres des signaux tatoués (en pointillés) et non tatoués (trait plein).

représentées par des symboles correspondant à différentes cyclo-fréquences, l'ensemble desquelles constitue la clé publique. La clé privée, qui permet de retrouver à la fois les données privées et publiques, consiste en l'ensemble de signaux cyclostationnaires. Des simulations préliminaires ont montré que des débits binaires allant de quelques dizaines à une centaine de bits par seconde peuvent être atteints avec des taux d'erreur raisonnables.

BIBLIOGRAPHIE

[1] L. Boney, A. Tewfik and K. Hamdy, « Digital watermarks for audio signals », *IEEE Int. Conf. on Multimedia Computing Systems*, Hiroshima, June 1996.

[2] L. Boney, A. Tewfik and K. Hamdy, « Digital watermarks for audio signals », *Eurospeech*, Rhodes, September 1997.

[3] E. Zwicker and E. Feldtkeller, « Psychoacoustique, l'oreille récepteur d'information », *Collection Technique et Scientifique des Télécommunications*, Masson, 1981.

[4] M. Perreau Guimarães, « Optimisation de l'allocation des ressources binaires et modélisation psychoacoustique pour le codage audio », Thèse de Doctorat, Université Paris V, Paris, 1998.

[5] T. Furon and P. Duhamel, « An asymmetric public detection watermarking technique », *Proc. of the 3rd Int. Work. on Information Hiding*, Dresden, September 1999.

[6] T. Furon, N. Moreau and P. Duhamel, « Audio public key watermarking technique », *Proc. Int. Conf. Acoust., Speech and Signal Processing*, Istambul, June 2000.

[7] J. Lacy, S. Quackenbush, A. Reibman, D. Shur and J. Snyder, « On combining watermarking with perceptual coding », *Proc. Int. Conf. Acoust., Speech and Signal Processing*, Seattle, May 1998.

[8] C. Neubauer and J. Herre, « Audio watermarking of MPEG-2 AAC bit streams scrambling », *108th AES Convention*, Paris, February 2000.

[9] E. Allamanche and J. Herre, « Secure delivery of compressed audio by compatible bitstream », *108th AES Convention*, Paris, February 2000.

[10] R.A. Garcia, « Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory », *107th AES Convention*, New York, September 1999.

[11] W. Bender, D. Gruhl, N. Morimoto and A. Lu, « Techniques for data hiding », *IBM System Journal*, Vol. 35, pp. 313-336, 1996.

[12] L. de C.T. Gomes, E. Gómez, M. Bonnet and N. Moreau, « Méthodes de resynchronisation pour le tatouage audio », *GRETSI 2001*, Toulouse, septembre 2001.

[13] N. Moreau, P. Dymarski, L. de C.T. Gomes, « Tatouage audio : une réponse à une attaque désynchronisante », *CORESA 2000*, Poitiers, octobre 2000.

[14] W.A. Gardner (ed.), « Cyclostationarity in communications and signal processing », *IEEE Press*, 1994.

[15] A. Dandawat and G.B. Giannakis, « Statistical tests for presence of cyclostationarity », *IEEE Transactions on Signal Processing*, Vol. 42, No. 9, September 1994.

Manuscrit reçu le 18 janvier 2001

LES AUTEURS

Mamadou MBOUP

Mamadou MBOUP est maître de conférences à l'UFR de mathématiques et informatique de l'université René Descartes-Paris V. Ses recherches en traitement de signal ont concerné ces dernières années l'identification de systèmes, l'égalisation autodidacte, le tatouage audio.
mboup@math-info.univ-paris5.fr

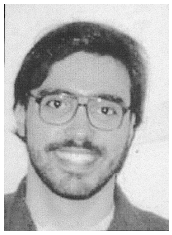
Madeleine BONNET

Madeleine BONNET est professeur à l'UFR de mathématiques et informatique de l'université René Descartes-Paris V. Ses recherches en traitement de signal ont concerné ces dernières années l'annulation d'écho acoustique, le codage de la parole et de la musique, le tatouage audio.
bonnet@math-info.univ-paris5.fr

Nicolas MOREAU

Nicolas MOREAU est diplômé de l'Institut National Polytechnique de Grenoble (IEG 1969). Il a obtenu une Habilitation à Diriger des Recherches en 1997 auprès de l'Université René Descartes-Paris V. Il est enseignant/chercheur à l'École Nationale Supérieure des Télécommunications. Ses activités de recherche concernent le traitement du signal dans le domaine du son (parole et musique) pour des applications de type multimédia. Plus précisément, elles portent sur les systèmes de compression, le tatouage, l'indexation, la spatialisation de sources sonores.
moreau@tsi.enst.fr

Leandro de C. T. GOMES



Leandro de C. T. GOMES, ingénieur en Génie Électrique diplômé de l'université de Campinas (Brésil), prépare actuellement une thèse à l'université René Descartes sur le tatouage audio.
tgomes@math-info.univ-paris5.fr