

Une nouvelle technique de tatouage audio par déplacement de phase des signaux à porteuses multiples

Mohammed KHALIL, Abdellah ADIB

Laboratoire Informatique de Mohammedia
B.P. 146 Mohammedia 20650, Maroc
medkhalil87@gmail.com et adib@fstm.ac.ma

Résumé – Cet article présente une nouvelle méthode de tatouage audio basée sur la modulation PSK (Phase Shift Keying). Cette technique, largement utilisée en télécommunications, a connu moins d'attention en tatouage audio. Cette méthode est mise en oeuvre par la régulation spécifique de certains paramètres du signal modulé. L'amplitude du signal modulé est contrôlée par le seuil de masquage afin de garantir l'inaudibilité du tatouage. Ensuite, l'information cachée est transmise dans la phase du signal modulé. Enfin, le choix judicieux de la fréquence porteuse conduit à introduire le concept d'accès multiple par répartition en fréquence FDMA (Frequency Division Multiple Access) afin d'augmenter la capacité d'information insérée.

Abstract – This paper presents a new audio watermarking method based on PSK (Phase Shift Keying) modulation. While there has been a considerable amount of attention devoted to this technique in telecommunications, there has only been a limited study for sinusoidal carrier watermarking. This method is implemented by a specific control of different modulated signal parameters. The amplitude of the modulated signal is controlled by the masking threshold to ensure the inaudibility constraint. Then, the hidden data is transmitted in the phase of the modulated signal. Finally, the judicious choice of the carrier frequency led us to introduce the concept of FDMA (Frequency Division Multiple Access) to increase the capacity of embedded information.

1 Introduction

Le tatouage audio, initialement orienté vers la protection des droits d'auteurs, peut être utilisé aussi pour la transmission de données [3]. Il s'agit de considérer le système de tatouage comme une chaîne de communication numérique aux propriétés très particulières [6][10]. Le signal porteur de l'information (le tatouage), est caractérisé par une faible puissance devant celle du bruit (le signal audio) dont les propriétés sont très différentes du bruit blanc gaussien (signal audio étant coloré et non stationnaire).

Dans ces dix dernières années, plusieurs techniques de tatouage pour la transmission de données ont été proposées. Dans [10], le système de tatouage utilise l'égaliseur de Wiener afin de perfectionner la détection du tatouage à la réception. Dans [3], les auteurs proposent des techniques d'insertions informées qui assurent la contrainte d'inaudibilité avant de maximiser la robustesse du tatouage face aux différentes perturbations externes. Un autre modèle est présenté dans [6], qui se base sur une nouvelle technique hybride qui regroupe la décomposition EMD (Empirical Mode Decomposition) et la technique de séparation de source ISA (Independent Subspace Analysis). Toutes ces approches utilisent la modulation par étalement de spectre à séquence directe dans leurs systèmes. Ce choix découle de deux principales raisons :

- Transmission du signal de tatouage avec une puissance faible et donc une inaudibilité assurée.

- Robustesse contre les interférences entre symboles et donc une bonne fiabilité de transmission.

En contrepartie, les modulations à bande étroite, largement utilisées en télécommunications, ont connu moins d'attention en tatouage audio. La principale cause est le choix judicieux de la fréquence porteuse f_p et de l'amplitude A du signal modulé qui permettent de satisfaire la contrainte d'inaudibilité. Toutefois, les modulations à bande étroite présentent quelques avantages, connus en communication numérique [7], que nous pouvons exploiter en tatouage audio. En effet, on se propose dans ce document de :

- Utiliser la modulation par déplacement de phase PSK¹ parcequ'elle permet une séparation de phase maximale et donc une grande robustesse vis-à-vis des interférences entre symboles [7].
- Ajuster l'amplitude du signal modulé pour satisfaire la contrainte d'inaudibilité.
- Exploiter la largeur de bande du signal audio en insérant plusieurs tatouages et ce, en se basant sur la technique FDMA².

Le reste du papier est organisé comme suit. La section 2 présente le système de tatouage par étalement de spectre. La technique de tatouage proposée fait l'objet de la section 3. La section 4 présente les résultats expérimentaux. Nous présenterons pour finir une conclusion qui fera l'objet de la section 5.

1. Phase Shift Keying.

2. Frequency Division Multiple Access.

2 Tatouage audio par étalement de spectre

L'information contenue dans le tatouage est représentée sous forme d'une séquence binaire b_i indépendante et identiquement distribuée. Elle est soumise dans un premier temps, à une phase de codage pour être convertie en une suite de symbole a_k prenant ses valeurs dans un alphabet de M symboles (chaque symbole a_k est le groupement de n bits codés). Le signal modulé $v(t)$ est construit en réalisant une association bijective entre les symboles a_k et les formes d'ondes (séquences pseudo-aléatoires) appartenant au dictionnaire de modulation. Ce signal ne peut pas être introduit directement dans le signal audio $x(t)$ du fait que sa puissance σ_v^2 est très faible par rapport à celle de $x(t)$. Afin d'augmenter sa puissance, tout en assurant la contrainte d'inaudibilité [3] [10], on le fait passer via un filtre de mise en forme perceptif $H(f)$. Le signal résultant $w(t)$ est ajusté ensuite, par un facteur d'échelle α pour assurer un compromis entre l'inaudibilité et la robustesse de la marque insérée. Enfin, le signal tatoué $y(t)$ est donné par :

$$\begin{aligned} y(t) &= \alpha w(t) + x(t) \\ &= \alpha v(t) * h(t) + x(t) \end{aligned} \quad (1)$$

où $h(t)$ est la réponse impulsionnelle du filtre $H(f)$ et $*$ est l'opération de convolution.

À la réception, le signal tatoué $y(t)$ est soumis à un annuleur d'interférences $\hat{H}^{-1}(f)$ qui non seulement inverse la réponse du canal, mais aussi décorrèle partiellement le bruit additif (signal audio). Ce filtre est estimé au récepteur à partir des propriétés de masquage du signal tatoué en utilisant un modèle psychoacoustique (MPA) au niveau du récepteur. Le signal estimé $\hat{v}(t)$ est soumis initialement à une étape de démodulation qui nécessite la connaissance des formes d'ondes utilisées à l'émetteur. Enfin, une étape de décision détermine la séquence binaire estimée \hat{b}_i .

3 Méthode de tatouage proposée

3.1 Modulation PSK adaptée au système de tatouage audio

La contrainte d'inaudibilité est assurée par la conception du filtre $H(f)$. Sa conception impose seulement de choisir un signal modulé $v(t)$ qui vérifie l'inégalité suivante [2] :

$$\sigma_v^2 \leq 1 \quad (2)$$

Si on utilise une modulation par étalement de spectre, le signal modulé $v(t)$ est blanc et de puissance unité. Toutefois, en utilisant la modulation PSK, les points de la constellation choisis sont habituellement placés avec un espacement angulaire uniforme autour d'un cercle [7]. Par conséquent, les signaux qui constituent le signal modulé $v(t)$ sont sinusoidaux avec une phase aléatoire [7][9] :

$$u(t, \varphi) = A \cos(2\pi f_p t + \varphi) \quad (3)$$

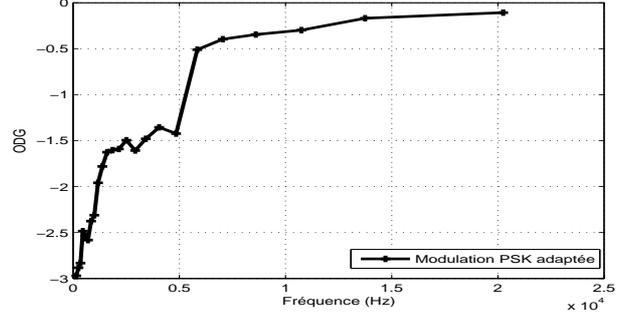


FIGURE 1 – Évaluation de la transparence du tatouage : ODG en fonction de f_p .

où la phase φ est une variable aléatoire qui est uniformément distribuée sur l'intervalle $[-\pi, \pi]$ avec une densité de probabilité [7] :

$$p(\varphi) = \frac{1}{2\pi} \quad \text{avec } \varphi \in [-\pi, \pi] \quad (4)$$

Afin d'utiliser le filtre $H(f)$, la puissance du signal modulé PSK doit satisfaire (2). D'après (3) et (4), on calcule les moments statistiques de la variable aléatoire $u(t_k, \varphi)$ à un instant donné t_k .

$$\begin{aligned} E[u(t_k, \varphi)] &= E[u(t_k)] \\ &= \int_{-\infty}^{\infty} p(\varphi) A \cos(2\pi f_p t_k + \varphi) d\varphi \\ &= \frac{1}{2\pi} A \int_{-\pi}^{\pi} \cos(2\pi f_p t_k + \varphi) d\varphi \\ &= 0 \end{aligned} \quad (5)$$

et

$$\begin{aligned} E[u^2(t_k, \varphi)] &= E[u^2(t_k)] \\ &= \int_{-\pi}^{\pi} p(\varphi) A^2 \cos^2(2\pi f_p t_k + \varphi) d\varphi \\ &= \frac{A^2}{2} \end{aligned} \quad (6)$$

Par conséquent, le signal modulé $v(t)$ est à moyenne nulle et de variance

$$\sigma_v^2 = \frac{A^2}{2} \quad (7)$$

Pour garantir la contrainte d'inaudibilité (2), on doit choisir

$$A \leq \sqrt{2} \quad (8)$$

3.2 Tatouage audio multiple basé sur la technique FDMA

La fréquence porteuse f_p est un paramètre important qui peut influencer sur la transparence du message inséré. En effet, après une étude préliminaire de l'impact des fréquences porteuses sur la transparence du tatouage, on a constaté que l'insertion dans les hautes fréquences permet de garantir mieux l'inaudibilité que dans les basses fréquences comme c'est illustré à la FIGURE 1. Ceci est dû à plusieurs raisons parmi lesquelles :

- Les études psychoacoustiques montrent que l'oreille est nettement moins sensible aux limites du spectre audible [5].

- La distribution énergétique du signal audio est concentrée autour des basses fréquences (au moins dans les signaux audio tests³).

Le fait d'introduire le tatouage dans les hautes fréquences sans presque aucune dégradation auditive ($ODG^4 \leq -0.5$), nous a conduit à insérer plus d'une seule information dans le même signal audio. En effet, le nouveau système de tatouage peut être vu comme un système de communication multi-utilisateurs, où les différents utilisateurs représentent les différents tatouages qui partagent tous le même canal qui est le signal audio [8]. La technique adoptée dans ce travail est FDMA qui permet de diviser la bande de fréquence du signal audio en plusieurs sous bandes. Un symbole du tatouage k est émis avec un signal sinusoïdal $u^{(k)}(t)$:

$$u^{(k)}(t) = \sqrt{2} \cos(2\pi(f_p + c^{(k)}\Delta f)t) \quad (9)$$

où f_p est la fréquence porteuse, Δf est l'écart de fréquence entre les tatouages adjacents et $c^{(k)}$ est le code FDMA assigné au tatouage k . Ainsi, le signal $v^{(k)}(t)$ assigné au tatouage k sera donné par :

$$v^{(k)}(t) = a^{(k)}u^{(k)}(t), \quad 0 \leq t < T_s \quad (10)$$

pour une durée de symbole T_s où $a^{(k)}$ représente le symbole transmis relative au tatouage k , $k = 0, \dots, K-1$ et K le nombre total de tatouages. Le signal modulé résultant de l'ensemble des K tatouages est donné par :

$$v(t) = \sum_{k=0}^{K-1} v^{(k)}(t) \quad (11)$$

4 Résultats expérimentaux

Douze signaux audio de différents styles (musique instrumentale, voix chantée et parole), échantillonnés à $F_e = 44.1$ kHz et apparentés à la base SQAM ; ont été utilisés dans tous les tests. Les performances du système ont été évaluées selon quatre critères : inaudibilité, fiabilité de transmission, capacité d'insertion et robustesse.

Pour évaluer l'inaudibilité du tatouage, on s'est servi d'une méthode de mesure objective basée sur l'algorithme PEAQ (Perceptual of Quality of Audio Perceived) qui fournit une valeur ODG variant de imperceptible (lorsque l'ODG est égal à 0) à très gênant (quand l'ODG est égal à -4) [1]. La TABLE 1 présente les valeurs ODG obtenues pour les deux systèmes : Le premier utilise l'étalement de spectre [4], tandis que le second utilise la modulation PSK adaptée avec $f_p = 5850$ Hz. La moyenne des valeurs ODG obtenues montre que la transparence du signal de tatouage est améliorée grâce à la modulation PSK adaptée présentée (Section 3.1). De plus, on remarque que les valeurs ODG minimales et maximales ont augmenté. Ceci prouve que la qualité auditive du signal tatoué est assurée pour chaque signal audio appartenant aux signaux tests.

3. Appartenant à la base de données SQAM.

4. Plus la valeur ODG est grande plus la qualité du signal tatoué est bonne

TABLE 1 – La valeur ODG du signal tatoué dans le cas d'utilisation de l'étalement de spectre et de la PSK

ODG	min	max	mean
Étalement de spectre	-3.6242	-0.6444	-1.3937
PSK adaptée	-0.7522	-0.1445	-0.5110

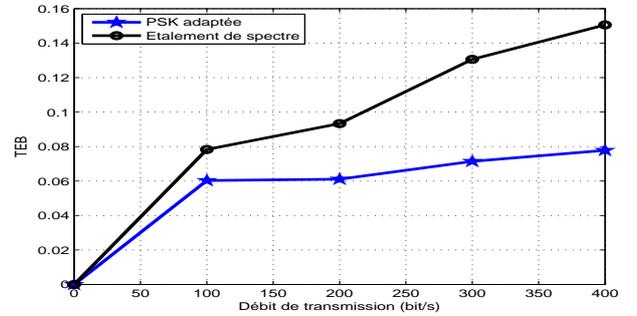


FIGURE 2 – Fiabilité de transmission.

Une fois, la transparence du message inséré est assurée, il serait nécessaire de le récupérer à partir du signal tatoué seul. La FIGURE 2 montre le gain apporté, en termes de TEB (Taux d'erreur Binaire), de la technique PSK adaptée par rapport à la technique d'étalement de spectre⁵ présentée dans [4].

Pour augmenter la capacité d'insertion du système de tatouage, on a opté pour utiliser la technique FDMA présentée dans la section 3.2. Deux facteurs limitent la capacité d'insertion : l'inaudibilité et la fiabilité de transmission. Pour ce, on a augmenté le nombre de tatouage insérés jusqu'à un certain degré défini par les valeurs ODG et TEB. Ensuite, on a comparé le nouveau système basé sur la technique FDMA avec celui basé sur la technique CDMA⁶ présenté dans [8].

D'après la FIGURE 3, on remarque que la technique FDMA assure mieux l'inaudibilité que la technique CDMA même quand on insère quatre tatouages simultanément. De plus, la technique FDMA permet une lente dégradation de la qualité du signal tatoué, lorsqu'on insère plus de tatouages, grâce au bon choix des bandes de fréquences attribuées aux différents tatouages. Ceci nous conduit à multiplier le signal de tatouage $w(t)$ par un facteur d'amplitude α variable afin d'augmenter sa puissance et faciliter ainsi sa détection.

D'après la FIGURE 4, on remarque que la technique CDMA permet une bonne fiabilité de transmission par rapport à la technique FDMA. En augmentant la valeur de α , on arrive à des performances équivalentes voir meilleures que CDMA. Cette fois ci, la technique CDMA permet une lente dégradation du TEB lorsqu'on insère plus de tatouages. Ceci est dû aux codes Walsh utilisés qui réduisent l'interférence d'accès multiple. Notons que chaque tatouage a été transmis avec un débit $R = 300$ bit/s, et donc on peut constater que la méthode proposée

5. Direct Sequence Spread Spectrum.

6. Code Division Multiple Access.

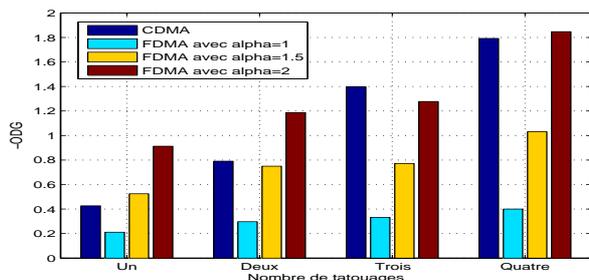


FIGURE 3 – ODG en fonction du nombre de tatouages insérés.

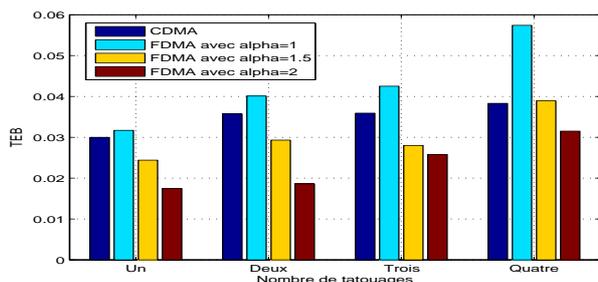


FIGURE 4 – TEB en fonction du nombre de tatouages insérés.

Perturbations	100 bit/s		200 bit/s	
	CDMA	FDMA	CDMA	FDMA
Filtre passe-haut	0.0042	0.0036	0.0250	0.0228
Filtre passe-bas	0.4175	0.2475	0.4758	0.3233
MPEG 96kbit/s	0.2894	0.0192	0.3183	0.0733
MPEG 64kbit/s	0.4658	0.0442	0.4967	0.2346

TABLE 2 – Robustesse du système de tatouage audio en termes de TEB, en utilisant les techniques CDMA et FDMA, contre différentes perturbations.

nous a permis de transmettre l’information avec un débit de $R = 1200\text{bit/s}$ et un $TEB = 10^{-2}$.

Enfin, le système de tatouage proposé est soumis à quatre types de perturbations : Filtre passe-haut, filtre passe-bas, compression MPEG 96 kbit/s et compression MPEG 64 kbit/s. La robustesse du système contre ces différentes perturbations, avec les débits $R = 100\text{bit/s}$ et $R = 200\text{bit/s}$, est présentée dans la TABLE 2.

On remarque que la technique FDMA est plus robuste au filtre passe-bas et au filtre passe-haut que la technique CDMA. Ceci est justifié par l’avantage que présente FDMA de ne pas utiliser certaines fréquences, si nécessaire. Par exemple les basses fréquences n’ont pas été utilisées lorsque le filtre passe-haut a été appliqué. La technique FDMA est robuste également contre la compression MPEG 96 Kbit/s et MPEG 64 Kbit/s parce que le choix des basses fréquences permet aux signaux de sauter certaines bandes qui présentent un évanouissement sélectif en fréquence.

5 Conclusion

Dans ce travail, nous avons présenté une nouvelle technique de tatouage basée sur le déplacement de phase des signaux à porteuses multiples. Les résultats expérimentaux ont montré quelques avantages d’utilisation des signaux à fréquences porteuses par rapport à ceux à spectre étalé. Une particularité de cette nouvelle méthode est que l’on peut choisir les fréquences convenables pour assurer la contrainte d’inaudibilité, éviter les sources connues d’interférences et enfin augmenter la capacité d’insertion.

Références

- [1] Recommendation B.S. 1387 : Method for objective measurements of perceived audio quality. *Int. Telecommunication Union, Geneva*, 2001.
- [2] C. Baras, N. Moreau, and P. Dymarski. Comparative study of two informed embedding strategies for audio spread-spectrum data hiding systems. *Proceedings of European Signal Processing Conference (EUSIPCO)*, 2005.
- [3] C. Baras, N. Moreau, and P. Dymarski. Controlling the inaudibility and maximizing the robustness in an audio annotation watermarking system. *IEEE Trans on Audio, Speech, and Language Processing*, vol. 14, no. 5, September 2006.
- [4] N. Cvejic, A. Keskinarkaus, and T. Seppanen. Audio watermarking using m-sequences and temporal masking. *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 227 – 230, October 2001.
- [5] O. Derrien, Sonia Larbi, Marcos Perreau Guimares, and Nicolas Moreau. Le codeur mpeg-2 aac expliqué aux traités de signaux. *Annales Des Télécommunications (Springer)*, 55(442-461), 2000.
- [6] N. El Hamdouni, A. Adib, S. Larbi, and M. Turki. A blind digital audio watermarking scheme based on EMD and UISA techniques. *Journal of Multimedia Tools and Applications, Springer*, 2011.
- [7] S. Haykin. *Digital Communications*. John Wiley and Sons, 4th edition 2001.
- [8] M. Khalil, N. El Hamdouni, and A. Adib. Increasing information capacity and improving detection reliability in audio watermarking system. In *ISCCSP, Rome, Italy*, 2012.
- [9] S. K. Kumarand and T. Sreenivas. Increased watermark-to-host correlation of uniform random phase watermarks in audio signals. In : *Signal Processing*, 87 (1). pp. 61-67, 2007.
- [10] S. Larbi, M. Jaidane, and N. Moreau. A new wiener filtering based detection scheme for time domain perceptual audio. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Montréal*, 2004.