



# THÈSE

Présentée en vue de l'obtention du  
Diplôme de Doctorat ès Sciences

*Filière* : Informatique

*Spécialité* : Réseaux et Systèmes d'Information Multimédia

Par :

**EUSCHI Salah**

---

## Une Approche de Tatouage Numérique Irréversible pour la Protection des Fichiers Audio

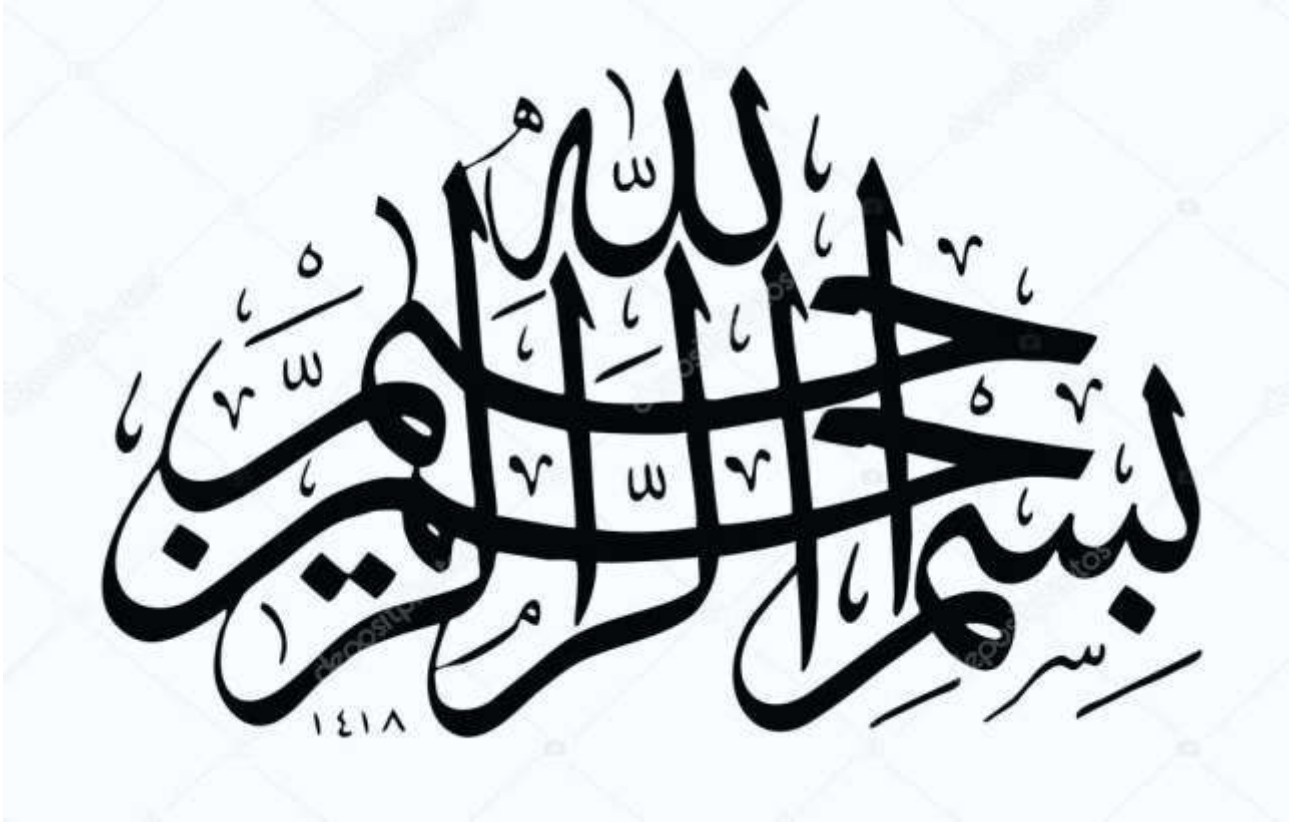
---

Soutenu publiquement le 09/11/2022

devant le jury composé de :

<i>Ahmed KORICHI</i>	<i>Professeur</i>	<i>Université de Ouargla</i>	<i>Président</i>
<i>Amine KHALDI</i>	<i>M.C.A</i>	<i>Université de Ouargla</i>	<i>Rapporteur</i>
<i>Mohammed. R. KAFI</i>	<i>M.C.A</i>	<i>Université de Ouargla</i>	<i>Co-Encadreur</i>
<i>Chaker Abdelaziz KERRACHE</i>	<i>M.C.A</i>	<i>Université de Laghouat</i>	<i>Examineur</i>
<i>Younes GUELLOUMA</i>	<i>M.C.A</i>	<i>Université de Laghouat</i>	<i>Examineur</i>
<i>Houcine BELLOUAAR</i>	<i>M.C.A</i>	<i>Université de Biskra</i>	<i>Examineur</i>

*Année universitaire : 2022/2023*



À ma petite fille Mélissa

## Remerciements

---

Je souhaite remercier en premier lieu mon directeur de thèse, M. Amine Khaldi, Maitre de conférences grade A , à l'Université Kasdi Merbah de Ouargla et Chef de département adjoint chargé de la post-graduation au département d'Informatique et des Technologies de l'Information, pour m'avoir encadré et accueilli au sein de son équipe de recherche. Je lui suis également reconnaissant pour le temps conséquent qu'il m'a accordé, ses qualités pédagogiques et scientifiques, sa franchise et sa sympathie. J'ai beaucoup appris à ses côtés et je lui adresse ma gratitude pour tout cela. J'adresse de chaleureux remerciements à mon Co-encadrant de thèse, M. Mohamed Redouane Kafi, Maitre de conférences grade A et Vice-recteur chargé de la post-graduation à l'université Kasdi Merbah de Ouargla, pour son attention de tout instant sur mes travaux, pour ses conseils avisés et son écoute qui ont été prépondérants pour la bonne réussite de cette thèse. Son énergie et sa confiance ont été des éléments moteurs pour moi. J'ai pris un grand plaisir à travailler avec lui.

J'adresse de sincères remerciements à M. Ahmed Korichi, Professeur à l'université de Ouargla et Président du conseil scientifique du département d'informatique et des Technologies de l'Information, pour avoir accepté de présider le jury de cette thèse. Ses remarques, son ouverture d'esprit, sa franchise, sa gentillesse sont autant d'éléments qui m'ont permis d'atteindre les objectifs demandés dans le cadre du doctorat. Je le remercie pour tout cela.

J'associe à ces remerciements M. Houcine Belouaar, Maitre de conférences grade A à l'université de Biskra, M. Chaker Abdelaziz Kerrache et M. Younes Guellouma, Maitres de Conférences Grade A à l'université de Laghouat, pour avoir accepté d'examiner mon travail.

Je tiens à remercier tous les membres, enseignants chercheurs et fonctionnaires du département d'informatique et des technologies de l'information, pour leur aide et leur bonne humeur. Nous avons partagé de bons moments. Je souhaite remercier spécialement Rachid Naimi pour son soutien et sa patience tout au long de la thèse.

Un grand merci à tous les enseignants chercheurs membres de notre équipe de recherche, M. Fares Kahlesnane, Maitre Assistant à l'université Kasdi Merbah Ouargla, et Mme Narima Zermi, Maitre de conférences grade A à l'université Bordj Badji Mokhtar Annaba.

## Résumé

---

L'objectif principal de cette thèse est la conception et l'implémentation d'approches de tatouage audio numérique aveugles. Ces approches doivent être efficaces, robustes et imperceptibles et sont destinées à protéger des fichiers audio, échangés en télé-médecine. Ce tatouage est réalisé par la dissimulation d'informations secrètes dans le fichier audio hôte et la marque insérée par un tel schéma de tatouage ne doit pas être effaçable, sans avoir recours à la clé ou à la méthode d'insertion. La marque est composée des informations du patient et des données d'acquisition. Ainsi, lors du processus d'extraction, une comparaison entre les informations du patient et la marque indiquera si les données ont été modifiées ou non. Cela permettra de s'assurer que l'échantillon correspond bien au patient indiqué dans l'enregistrement. Trois contributions originales ont été proposées dans cette thèse.

Dans notre première contribution intitulée " *Un algorithme de tatouage audio basé sur la transformée de Fourier pour la protection des sons respiratoires* "; nous avons proposé deux nouveaux schémas de substitution pour le tatouage audio numérique basé sur la transformée de Fourier. L'insertion de la marque a été réalisée en combinant la parité des valeurs des coefficients successifs, chaque variante représente une combinaison différente. Dans nos expériences, plusieurs variantes de la transformée de Fourier sont utilisées à savoir une Transformée de Fourier Discrète, une Transformée de Fourier Fractionnelle ainsi qu'une Transformée de Fourier Discrète Quaternion. Pour chaque transformée, les deux variantes de notre schéma de tatouage sont appliquées et les résultats obtenus ont montré que notre approche offre une bonne imperceptibilité et génère un échantillon audio tatoué robuste contre diverses attaques avec une marque de haute qualité. Cependant, l'utilisation de petits fichiers audio pour nos expériences réduit considérablement la capacité de notre approche dans le domaine fréquentiel.

Dans notre deuxième contribution intitulée " *Approches spatiales et fréquentielles pour la protection des fichiers audio* "; nous avons proposé un nouveau schéma d'insertion pour le tatouage aveugle des fichiers audio. L'objectif est de trouver un compromis entre la capacité et l'imperceptibilité afin de cacher autant de données que possible tout en minimisant la dégradation du fichier. Ce schéma d'insertion est mis en œuvre dans les trois domaines d'insertion : spatial, fréquentiel et multi-résolution. Pour l'insertion dans le domaine spatial, la marque est insérée directement dans les échantillons de données. Pour l'insertion dans le domaine fréquentiel, une transformée en cosinus discrète est appliquée aux trames audio, après l'étape de seuillage et de quantification, la marque est insérée dans les coefficients de la transformée en cosinus discrète pour obtenir le fichier tatoué. Pour l'insertion dans le domaine multi-résolution, une transformée en ondelettes discrètes à un seul niveau est appliquée en utilisant le filtre passe-bas de mise à l'échelle et le filtre passe-haut d'ondelettes. L'insertion de la marque est ensuite effectuée à l'aide des coefficients AC obtenus. Le processus de dissimulation proposé combine trois valeurs pour insérer deux bits et une seule valeur peut être modifiée, ce qui réduit la probabilité de modification contrairement aux autres approches. Cela implique moins de modification et donc moins de distorsion du fichier hôte, ce qui explique le bon rapport signal/bruit obtenu de plus de 59dB pour l'insertion dans le domaine spatial et donc une imperceptibilité raisonnable. Une évaluation de la robustesse du tatouage démontre que les schémas proposés génèrent des échantillons tatoués

---

raisonnablement robustes contre diverses attaques et une marque insérée de haute qualité avec une corrélation croisée normalisée supérieure à 0,9 pour les trois domaines d'insertion.

Dans notre troisième contribution intitulée " *Un schéma de tatouage audio basé sur la combinaison de valeur de parité pour la protection des sons de battements de cœur* " et afin de protéger les données échangées en télémédecine ; nous avons proposé un schéma de tatouage aveugle pour la protection des sons du rythme cardiaque. Deux schémas innovants de tatouage aveugle ont été proposés, chaque schéma élaboré a été décliné en deux variantes spécifiques. L'insertion réussie de la marque se fait typiquement en combinant trois valeurs successives de parité, chaque variante représente une combinaison différente. Ces approches ont été mises en œuvre dans les trois domaines d'insertion : spatial, fréquentiel et multi-résolution. Pour le domaine spatial, la marque est insérée dans les valeurs d'échantillons du fichier. Dans le domaine fréquentiel, les bits de la marque sont substitués au bit le moins significatif du coefficient DCT. Pour l'insertion dans le domaine multi-résolution, après avoir calculé une DWT à un seul niveau, les coefficients de sous-bande AC obtenus sont utilisés pour le processus d'insertion. Après avoir comparé les approches que nous avons utilisées aux divers travaux récents dans les trois domaines utilisés, les résultats obtenus démontrent suffisamment que notre deuxième approche proposée offre une bonne imperceptibilité pour l'insertion observée dans les domaines fréquentiel et spatial. Cependant, l'utilisation de petits fichiers audio pour nos expériences réussies réduit considérablement la capacité de nos méthodes utilisées dans le domaine fréquentiel.

*Descripteurs : Tatouage audio numérique, tatouage aveugle, bit le moins significatif, transformée de Fourier discrète, transformée en Cosinus discrète, transformée en ondelettes discrètes, domaine spatial, domaine de transformées.*

## Abstract

---

The main objective of this thesis is the design and implementation of blind digital audio watermarking approaches. These approaches must be efficient, robust and imperceptible and are intended to protect audio files, exchanged in telemedicine. This watermarking is achieved by hiding secret information in the host audio file and the mark inserted by such a watermarking scheme must not be erasable, without the use of the key or insertion method. The mark is composed of the patient information and the acquisition data. Thus, during the extraction process, a comparison between the patient information and the mark will indicate whether or not the data has been altered. This will ensure that the sample matches the patient indicated in the record. Three original contributions were proposed in this thesis.

In our first contribution entitled "*A Fourier transform-based audio watermarking algorithm for respiratory sound protection*"; we propose two new substitution schemes for digital audio watermarking based on the Fourier transform. The integration of the watermark will be performed by combining the parity of the successive coefficients values; each variant will represent a different combination. In our experiments several variants of the Fourier transform are used (Discrete Fourier Transform, a Fractional Fourier Transform as well as a Quaternion Discrete Fourier Transform). For each transform, both variants of our watermarking scheme are applied and the results obtained show that our approach offers good imperceptibility and generates watermarked audio sample robust against various attacks with a high quality watermark. However, using small audio file for our experiments considerably reduces the capacity of our approach in the frequency domain.

In our second contribution entitled "*Spatial and frequency approaches for audio file protection*" and in order to protect audio files, we propose a new integration scheme for blind audio file watermarking. The goal is to find a compromise between capacity and imperceptibility in order to hide as much data as possible while minimizing file degradation. This integration scheme is implemented in the three insertion domains: spatial, frequency and multi-resolution domain. For the spatial domain integration, the mark is inserted directly into the data samples. For the frequency domain integration, a Discrete Cosine Transform is applied to the audio frames, after the thresholding and quantification step the watermark is inserted into the Discrete Cosine Transform coefficients to obtain the watermarked file. For the multi-resolution domain insertion, a single level Discrete Wavelet Transform is applied using the scaling low pass filter and wavelet high pass filter. The watermark integration is then performed using the obtained AC coefficients. The proposed concealment process combines three values to integrate two bits and only one may be modified, which reduces the probability of change unlike other approaches. This implies less modification and therefore less distortion of the host file, this explains the good Signal to Noise Ratio obtained more than 59dB for the spatial domain integration and therefore a reasonable imperceptibility. An evaluation of the watermark's robustness demonstrates that the proposed schemas generate reasonably robust watermarked samples against various attacks with a high-quality watermark with a normalized cross correlation greater than 0.9 for the three insertion domain.

In our third contribution entitled "*A parity value combination based audio watermarking scheme for heartbeat sound protection*" and in order to enhance the security of data exchanged in telemedicine, we propose a blind watermarking scheme for heartbeat sounds protection. Two

innovative schemes of blind watermarks are proposed; each elaborate scheme will be declined into two specific variants. The successful watermark integration will be typically done by combining three successive values parity; each variant will adequately represent a different combination. These approaches will be implemented in the three insertion domains: spatial, frequency and multi-resolution domain. For the spatial domain, the watermark will be integrated into the samples values of the file. In the frequency domain, the watermark bits will be substituted to the DCT coefficient's least significant bit. For the multi-resolution domain insertion, after calculating a DWT, the obtained AC sub-band coefficients will be used for the integration process. After comparing our used approaches to the various recent works in the three used domains; the obtained results sufficiently demonstrate that our second proposed approach offers a good imperceptibility for the observed frequency and spatial domains insertion. However, using small audio file for our successful experiments significantly reduces the capacity of our used methods in the frequency domain.

*Keywords:* Digital audio watermarking, blind watermarking, least significant bit, discrete Fourier transform, discrete cosine transform, discrete wavelet transform, spatial domain, transform domain.

الهدف الرئيسي من هذه الأطروحة هو تصميم وتنفيذ مناهج العلامات الوسمية الصوتية الرقمية العمياء. يجب أن تكون هذه الأساليب فعالة وقوية وغير محسوسة وتهدف إلى حماية الملفات الصوتية التي يتم تبادلها في التطبيق عن بعد. يتم تحقيق هذه العلامة المائية عن طريق إخفاء المعلومات السرية في ملف الصوت المضيف ويجب ألا تكون العلامة التي تم إدخالها بواسطة مخطط العلامات المائية هذه قابلة للمسح دون استخدام المفتاح أو طريقة الإدراج. تتكون العلامة من معلومات المريض وبيانات الاستحواذ. وبالتالي، أثناء عملية الاستخراج، ستشير المقارنة بين معلومات المريض والعلامة إلى ما إذا كان قد تم تغيير البيانات أم لا. سيضمن ذلك تطابق العينة مع المريض المشار إليه في السجل. تم اقتراح ثلاث مساهمات أصلية في هذه الأطروحة.

في مساهمتنا الأولى بعنوان "خوارزمية للعلامات الوسمية الصوتية القائمة على التحويل من فورييه لحماية اصوات الجهاز التنفسي"؛ نقترح مخططين جديدين لاستبدال العلامة الوسمية الصوتية الرقمية بناءً على تحويل فورييه. سيتم تنفيذ إدراج العلامة الوسمية من خلال الجمع بين تكافؤ قيم المعاملات المتتالية؛ كل متغير سيمثل مجموعة مختلفة. في تجاربنا، تم استخدام العديد من المتغيرات لتحويل فورييه (تحويل فورييه المنفصل، تحويل فورييه الجزئي بالإضافة إلى تحويل فورييه المتقطع الرباعي). لكل تحويل، يتم تطبيق كلا المتغيرين من مخطط العلامات الوسمية لدينا وتظهر النتائج التي تم الحصول عليها أن نهجنا يوفر عدم إدراك جيد ويولد عينة صوتية تحمل علامة وسمية قوية ضد الهجمات المختلفة بعلامة وسمية عالية الجودة. ومع ذلك، فإن استخدام ملف صوتي صغير لتجاربنا يقلل بشكل كبير من قدرة نهجنا في مجال التردد.

في مساهمتنا الثانية بعنوان "المقاربات المكانية والترددية لحماية الملفات الصوتية" ومن أجل حماية الملفات الصوتية، نقترح مخطط تكامل جديد للعلامة الوسمية لملفات الصوت العمياء. الهدف هو إيجاد حل وسط بين السعة وعدم الإدراك من أجل إخفاء أكبر قدر ممكن من البيانات مع تقليل تدهور الملف. يتم تنفيذ مخطط التكامل في مجالات الإدراج الثلاثة: المجال المكاني والتردد ومجال الاستبانة المتعددة. للتكامل في المجال المكاني، يتم إدراج العلامة مباشرة في عينات البيانات. للإدراج في مجال التردد، يتم تطبيق تحويل جيب التمام المنفصل على الإطارات الصوتية، بعد خطوة العتبة والتقدير الكمي، يتم إدخال العلامة الوسمية في معاملات تحويل جيب التمام المنفصل للحصول على الملف ذي العلامة الوسمية. للإدراج في المجال متعدد الدقة، يتم تطبيق تحويل الموجة المنفصلة على مستوى واحد باستخدام مرشح التمرير المنخفض التحجيم ومرشح التمرير العالي المويج. ثم يتم تنفيذ إدراج العلامة الوسمية باستخدام معاملات التيار المتردد التي تم الحصول عليها. تجمع عملية الإخفاء المقترحة بين ثلاث قيم لدمج بيتين ويمكن تعديل قيمة واحدة فقط، مما يقلل من احتمالية التغيير على عكس الأساليب الأخرى. هذا يعني تعديلاً أقل وبالتالي تشويهاً أقل لملف المضيف، وهذا يفسر نسبة الإشارة إلى الضوضاء الجيدة التي تم الحصول عليها أكثر من 59 ديسيبل لتكامل المجال المكاني وبالتالي عدم الإدراك المعقول. يوضح تقييم متانة العلامة الوسمية أن المخططات المقترحة تولد عينات قوية بشكل معقول من العلامات الوسمية ضد الهجمات المختلفة بعلامة وسمية عالية الجودة مع ارتباط متقاطع طبيعي أكبر من 0.9 لمجالات الإدراج الثلاثة.

في مساهمتنا الثالثة بعنوان "مخطط العلامات الوسمية الصوتية على أساس مزيج قيم التكافؤ لحماية صوت ضربات القلب" ومن أجل تعزيز أمن البيانات المتبادلة في التطبيق عن بعد، نقترح مخطط علامة وسمية عمياء لحماية أصوات ضربات القلب. تم اقتراح خطتين مبتكرتين للعلامات الوسمية العمياء؛ سيتم وضع كل مخطط تفصيلي إلى متغيرين محددتين. عادةً ما يتم تكامل العلامة الوسمية الناجح من خلال الجمع بين ثلاث قيم متتالية التكافؤ؛ كل متغير سيمثل بشكل مناسب مجموعة مختلفة. سيتم تنفيذ هذه الأساليب في مجالات الإدراج الثلاثة: المجال المكاني والتردد ومجال الاستبانة المتعددة. بالنسبة للمجال المكاني، سيتم دمج العلامة الوسمية في عينات قيم الملف. في مجال التردد، سيتم استبدال بتات العلامة الوسمية بالبت الأقل أهمية لمعامل DCT بالنسبة لإدراج المجال متعدد الاستبانة، بعد حساب DWT، سيتم استخدام معاملات النطاق الفرعي AC التي تم الحصول عليها لعملية التكامل. بعد مقارنة مناهجنا المستخدمة مع الأعمال الحديثة المختلفة في المجالات الثلاثة المستخدمة؛ توضح النتائج التي تم الحصول عليها بشكل كافٍ أن نهجنا الثاني المقترح يوفر عدم إدراك جيد للتردد المرصود والمجالات المكانية. ومع ذلك، فإن استخدام ملف صوتي صغير لتجاربنا الناجحة يقلل بشكل كبير من قدرة الأساليب المستخدمة لدينا في مجال التردد.



الكلمات المفتاحية: العلامات الرسمية الصوتية الرقمية ، العلامات الرسمية العمياء ، البيت الأقل أهمية ، تحويل فورييه المنفصل ، تحويل جيب التمام المنفصل ، التحويل المويجي المنفصل ، المجال المكاني ، مجال التحويل

## Principales abréviations

---

<b>AAC</b>	<i>Advanced Audio Coding</i>
<b>AC</b>	<i>Approximation Coefficients</i>
<b>AIFF</b>	<i>Audio Interchange File Format</i>
<b>ANOVA</b>	<i>Analysis of Variance</i>
<b>ATH</b>	<i>Absolute Threshold of Hearing</i>
<b>BER</b>	<i>Bit Error Rate</i>
<b>CORR</b>	<i>Correlation Coefficient</i>
<b>DC</b>	<i>Detail Coefficients</i>
<b>DCT</b>	<i>Discrete Cosine Transform</i>
<b>DFT</b>	<i>Discrete Fourier Transform</i>
<b>DWT</b>	<i>Discrete Wavelet Transform</i>
<b>FFT</b>	<i>Fast Fourier Transform</i>
<b>FrFT</b>	<i>Fractional Fourier transform</i>
<b>HAS</b>	<i>Human Auditory System</i>
<b>HVS</b>	<i>Human Visual System</i>
<b>IDCT</b>	<i>Inverse DCT</i>
<b>IDWT</b>	<i>Inverse DWT</i>
<b>IFPI</b>	<i>International Federation of Phonographic Industry</i>
<b>LSB</b>	<i>Low-Significant-Bit</i>
<b>LWT</b>	<i>Lifting Wavelet Transform</i>
<b>MNR</b>	<i>Mask to Noise Ratio</i>
<b>MOS</b>	<i>Mean Opinion Score</i>
<b>MP3</b>	<i>MPEG-1/2 Audio Layer 3</i>
<b>MSE</b>	<i>Mean Squared Error</i>
<b>NC</b>	<i>Normalized Correlation</i>
<b>ODG</b>	<i>Objective Difference Grade</i>
<b>PCM</b>	<i>Pulse Code Modulation</i>
<b>PEAQ</b>	<i>Perceptual Evaluation of Audio Quality</i>
<b>PSNR</b>	<i>Peak Signal to Noise Ratio</i>
<b>QDFT</b>	<i>Quaternion Discrete Fourier transform</i>

<b><i>QIM</i></b>	<i>Quantization Index Modulation</i>
<b><i>SDG</i></b>	<i>Subjective Difference Grade</i>
<b><i>SMR</i></b>	<i>Signal to Mask Ratio</i>
<b><i>SNR</i></b>	<i>Signal to Noise Ratio</i>
<b><i>SPL</i></b>	<i>Sound Pressure Level</i>
<b><i>SS</i></b>	<i>Spread Spectrum</i>
<b><i>SSIM</i></b>	<i>Structural Similarity Index</i>
<b><i>SSM</i></b>	<i>Spread Spectrum based Modulation</i>
<b><i>SVD</i></b>	<i>Singular Value Decomposition</i>
<b><i>WAVE</i></b>	<i>Waveform audio format</i>

# Table des matières

---

<b>Remerciements</b>	<b><i>i</i></b>
<b>Résumé</b>	<b><i>ii</i></b>
<b>Abstract</b>	<b><i>iv</i></b>
<b>نبذة مختصرة</b>	<b><i>vi</i></b>
<b>Principales abréviations</b>	<b><i>viii</i></b>
<b>Table des matières</b>	<b><i>x</i></b>
<b>Liste des figures</b>	<b><i>xviii</i></b>
<b>Liste des tableaux</b>	<b><i>xx</i></b>
<b>1. Introduction Générale</b>	<b><i>1</i></b>
<b>1.1 Les techniques de dissimulation de l'information</b>	<b><i>2</i></b>
1.1.1 La Cryptographie	<i>2</i>
1.1.2 La Stéganographie	<i>4</i>
1.1.3 Le Tatouage	<i>5</i>
<b>1.2 Un aperçu du tatouage numérique</b>	<b><i>6</i></b>
<b>1.3 Problématique</b>	<b><i>8</i></b>
<b>1.4 Organisation de la thèse</b>	<b><i>10</i></b>
<b>2. Le Son et le Traitement Numérique des Fichiers Audio</b>	<b><i>12</i></b>
<b>2.1 Introduction</b>	<b><i>13</i></b>
2.1.1 Définitions Média et Multimédia	<i>13</i>
2.1.2 Médias discrets / Médias continus	<i>14</i>
2.1.3 Caractéristiques des médias continus	<i>14</i>
2.1.4 Numérisation	<i>15</i>
<b>2.2 Le son numérique</b>	<b><i>15</i></b>
2.2.1 Les caractéristiques physiques du son	<i>15</i>
2.2.1.1 La nature du son	<i>15</i>
2.2.1.2 Caractéristiques du son	<i>16</i>

---

2.2.1.3	Analyse de Fourier et représentations spectrales	17
2.2.2	La mesure du niveau acoustique	19
2.2.3	Numérisation d'un signal audio analogique	20
2.2.3.1	L'échantillonnage du son	20
2.2.3.2	La quantification	21
2.2.3.3	Modulation par impulsion codée PCM (Pulse Code Modulation)	22
2.2.4	La compression du signal audio numérisé	23
2.2.4.1	Les algorithmes de compression	23
2.2.4.2	Le taux de compression	23
<b>2.3</b>	<b>Format de fichier audio</b>	<b>24</b>
2.3.1	Caractéristique des formats audio	24
2.3.2	Les formats audio sans compression	24
2.3.3	Les formats audio compressés sans perte	25
2.3.4	Les formats audio compressés avec perte	26
<b>2.4</b>	<b>Codage audio perceptif utilisant des modèles psycho-acoustiques</b>	<b>28</b>
2.4.1	Les classes du signal audio	28
2.4.2	Débit binaire requis pour l'audio de qualité stéréo	29
2.4.3	Portée des normes audio MPEG	30
2.4.4	Les technologies clés du codage audio	30
2.4.4.1	Un aperçu sur les techniques du codage perceptuel	31
2.4.4.2	La perception auditive humaine	33
1)	Le phénomène de masquage	33
2)	Masquage temporel	35
2.4.4.3	Codage perceptuel dans le MPEG audio	35
<b>2.5</b>	<b>Conclusion</b>	<b>36</b>
<b>3.</b>	<b><i>Etat de l'Art sur les Techniques du Tatouage Audio Numérique</i></b>	<b>38</b>
<b>3.1</b>	<b>Introduction</b>	<b>39</b>
<b>3.2</b>	<b>Concept du tatouage audio numérique</b>	<b>39</b>
3.2.1	Concepts de base et termes généraux	39
3.2.2	Schéma général du tatouage numérique	40
3.2.2.1	Processus d'insertion	40
1)	La génération de la marque	40
2)	L'insertion de la marque :	41
3.2.2.2	Processus d'extraction	41
3.2.2.3	Les types de marque et leur détection	42
1)	Séquence gaussienne pseudo-aléatoire	42
2)	Marques d'images binaires ou d'images en niveaux de gris	42
3.2.3	Les propriétés requises d'un système de tatouage audio numérique	43

---

3.2.3.1 Imperceptibilité	43
3.2.3.2 Robustesse	43
3.2.3.3 Capacité	44
3.2.3.4 Sécurité	44
3.2.3.5 Complexité et Coût de calcul	45
3.2.3.6 Détection aveugle ou non aveugle	45
<b>3.3 Classification des systèmes de tatouage audio numérique</b>	<b>45</b>
3.3.1 Selon le domaine de travail	46
3.3.1.1 Insertion dans le domaine spatial	46
3.3.1.2 Insertion dans le domaine de transformation ou fréquentiel	46
3.3.1.3 Insertion dans le domaine Cepstral	47
3.3.1.4 Algorithmes hybrides	47
3.3.2 Selon la perception humain	47
3.3.2.1 Tatouage robuste	48
3.3.2.2 Tatouage fragile	48
3.3.2.3 Tatouage semi-fragile	48
3.3.3 Selon la technique d'insertion	48
3.3.3.1 Systèmes de tatouage additifs	48
3.3.3.2 Systèmes de tatouage substitutifs	48
3.3.4 Selon le champ d'application	49
3.3.4.1 Protection des droits d'auteur	49
3.3.4.2 Authentification et contrôle d'intégrité	49
3.3.4.3 Transmission de données	49
3.3.4.4 Applications médicales	49
3.3.4.5 Autres applications	50
3.3.5 Selon le processus de détection	50
3.3.5.1 Tatouage non aveugle	50
3.3.5.2 Tatouage semi-aveugle	50
3.3.5.3 Tatouage aveugle	51
<b>3.4 Les techniques fondamentales de tatouage audio numérique</b>	<b>51</b>
3.4.1 Tatouage audio du domaine spatial	52
3.4.1.1 Tatouage Audio par Substitution LSB (Low-Significant-Bit Coding)	52
3.4.1.2 Tatouage audio additif	53
3.4.1.3 Masquage de l'écho	54
3.4.1.4 Codage de phase	54
3.4.1.5 Etalement de spectre	55
3.4.1.6 Algorithme de Patchwork	56
3.4.1.7 La modulation de l'indice de quantification (QIM- Quantization index modulation)	58
3.4.1.8 Tatouage numérique audio résistant et affine utilisant un modèle de synchronisation	59

3.4.1.9 Tatouage audio numérique basé sur le saut de bits multiples et le système auditif humain	59
3.4.1.10 Technique de tatouage audio basée sur les périodes de silence	60
3.4.2 Tatouage audio du domaine de transformation	60
3.4.2.1 Les transformées de Fourier	61
1) Transformée de Fourier discrète -DFT	61
2) Transformée de Fourier fractionnaire -FRFT	62
3) Transformée de Fourier discrète en quaternion-QDFT	62
4) Caractéristiques des transformées de Fourier	63
5) Critères de sélection des coefficients	63
6) Avantage de la DFT par rapport aux autres transformées	63
3.4.2.2 La Transformée en Cosinus Discrète – DCT	64
1) Propriétés de la DCT pour le traitement de signal	66
3.4.2.3 La Transformée en ondelette Discrète – DWT	66
1) Propriétés de la DWT par rapport au traitement de signal	69
2) Avantages de la DWT par rapport à la DCT	70
3) Inconvénients de la DWT par rapport à la DCT	70
3.4.2.4 Décomposition en valeur singulière (Singular Value Decomposition-SVD)	70
3.4.2.5 Tatouage audio par Substitution basé DCT	70
3.4.2.6 Tatouage Audio Multiplicatif basé DCT/DWT/SVD	71
3.4.2.7 Tatouage audio additif basé SVD	71
3.4.3 Tatouage audio de domaine Cepstral	72
3.4.4 Tatouage audio de domaine hybride	73
<b>3.5 Travaux connexes dans le domaine de tatouage numérique audio</b>	<b>75</b>
3.5.1 Méthodes de tatouage audio dans le domaine spatial	75
3.5.2 Méthodes de tatouage audio dans le domaine de transformation DFT / DCT	76
3.5.3 Méthodes de tatouage audio dans le domaine multi résolution DWT	77
<b>3.6 Conclusion</b>	<b>78</b>
<b>4. Critères de Mesure de Performance des Systèmes de Tatouage Audio</b>	
<b>Numérique</b>	<b>80</b>
<b>4.1 Introduction</b>	<b>81</b>
<b>4.2 Le système auditif humain</b>	<b>81</b>
4.2.1 Propriétés de l'audition humaine monaurale	81
4.2.2 Le seuil absolu d'audition (ATH -Absolute Threshold of Hearing)	83
4.2.3 Les bandes critiques	84
4.2.4 Le phénomène de masquage	85
4.2.5 Le masquage temporel	88

<b>4.3</b>	<b>Les attaques sur les systèmes de tatouage audio numérique</b>	<b>88</b>
4.3.1	Classification de Guang Hua et al.	90
4.3.1.1	Les attaques de base	91
4.3.1.2	Les attaques avancées	91
4.3.1.3	Effets des attaques sur la qualité de l'audio tatoué	92
4.3.2	Classification de Steinebach et al.	93
4.3.2.1	Test de Robustesse de Base	95
4.3.2.2	Test de robustesse avancé	98
4.3.3	Outils pour l'évaluation des schémas de tatouage audio	99
4.3.3.1	Certimark : Certification des techniques de tatouage	99
4.3.3.2	Checkmark	99
4.3.3.3	Test SDMI	100
4.3.3.4	STEP 2000	101
4.3.3.5	StirMark Benchmark	101
<b>4.4</b>	<b>Evaluation de l'imperceptibilité</b>	<b>102</b>
4.4.1	Evaluation subjective	103
4.4.1.1	Évaluation de la transparence : ABX	103
4.4.1.2	Évaluation de la qualité absolue : MOS (mean opinion score)	104
4.4.1.3	Évaluation des petites dépréciations : SDG (subjective difference grade)	105
4.4.2	Evaluation objective	106
4.4.2.1	SNR (Signal to Noise Ratio)	106
4.4.3	Evaluation objective perceptive	107
4.4.3.1	Le grade de différence objective : ODG (Objective Difference Grade)	107
4.4.3.2	L'algorithme PEAQ	107
4.4.3.3	L'imperceptibilité pour le contrôle des marques insérées	109
4.4.4	Les régions d'insertion des marques	109
4.4.5	Niveaux d'assurance possibles pour l'évaluation de l'imperceptibilité	111
<b>4.5</b>	<b>Evaluation de la robustesse</b>	<b>112</b>
4.5.1	Mesures mathématiques	112
4.5.2	Procédure de base pour le test de robustesse face aux attaques	114
<b>4.6</b>	<b>Evaluation de la fiabilité</b>	<b>115</b>
4.6.1	Les fausses alarmes	115
4.6.2	Niveaux d'assurance possibles pour l'évaluation de la fiabilité et la robustesse	115
<b>4.7</b>	<b>Evaluation de la capacité</b>	<b>116</b>
<b>4.8</b>	<b>Conclusion</b>	<b>117</b>
<b>5.</b>	<b><i>Nouvelles Approches de Tatouage Numérique pour la Protection des Fichiers Audio</i></b>	<b>119</b>



<b>5.1</b>	<b>Introduction</b>	<b>120</b>
<b>5.2</b>	<b>Un algorithme de tatouage numérique audio basé sur la transformée de Fourier pour la protection des sons respiratoires</b>	<b>120</b>
5.2.1	Contexte d'étude et problématique	120
5.2.2	Schéma de tatouage proposé	121
5.2.3	Résultats expérimentaux et tests	123
5.2.3.1	Test d'imperceptibilité	125
5.2.3.2	Test de robustesse	126
5.2.4	Analyse et discussions des résultats	126
<b>5.3</b>	<b>Approches spatiales et fréquentielles pour la protection des fichiers audio</b>	<b>127</b>
5.3.1	Contexte d'étude et problématique	127
5.3.2	Approches proposées pour le tatouage aveugle	128
5.3.2.1	Domaine spatial	128
5.3.2.2	Domaine fréquentiel	129
1)	Utilisation de la transformée en cosinus discrète (DCT)	129
	Processus de quantification	130
	Processus d'insertion	130
5.3.2.3	Domaine Multi-résolution DWT	130
5.3.3	Résultats expérimentaux	131
5.3.3.1	Mesures de la performance	131
5.3.3.2	Application de nos approches dans le domaine spatial	133
5.3.3.3	Application de nos approches dans le domaine fréquentiel	133
5.3.3.4	Application de nos approches dans le domaine de la multi-résolution	134
5.3.3.5	Comparaison des résultats	135
5.3.3.6	Test de robustesse	137
5.3.4	Analyse et discussion des résultats	137
<b>5.4</b>	<b>Un schéma de tatouage basé sur la combinaison de valeur de parité pour la protection des sons de battements de cœur</b>	<b>138</b>
5.4.1	Contexte d'étude et problématique	138
5.4.2	Approches proposées	139
5.4.2.1	Domaine spatial	139
5.4.2.2	Domaine fréquentiel	139
1)	La transformée en cosinus discrète (DCT)	140
	Processus de quantification	140
	Processus d'insertion	140
5.4.2.3	Domaine multi-résolution	141
5.4.3	Expériences et résultats	141
5.4.3.1	Mesures de la performance	142
5.4.3.2	Application de nos approches dans le domaine spatial	142

5.4.3.3 Application de nos approches dans le domaine fréquentiel .....	143
5.4.3.4 Application de nos approches dans le domaine de la multi-résolution .....	144
5.4.3.5 Test d'imperceptibilité .....	145
5.4.3.6 Test de robustesse .....	146
5.4.4 Analyse et discussion des résultats .....	146
<b>5.5 Conclusion et perspectives .....</b>	<b>147</b>
<b>6. Conclusion générale et perspectives .....</b>	<b>149</b>
<b>6.1 Rappels sur les contributions .....</b>	<b>150</b>
<b>6.2 Perspectives et travaux futurs .....</b>	<b>151</b>
<b><i>Bibliographie</i> .....</b>	<b>153</b>
<b><i>Nos publications</i> .....</b>	<b>168</b>
<b>Publications internationales .....</b>	<b>168</b>
<b>Conférences nationales .....</b>	<b>168</b>
<b>Publications internationales comme co-auteur .....</b>	<b>168</b>
<b><i>Table des annexes</i> .....</b>	<b>170</b>
<b><i>Annexe 1 – Base de données des sons respiratoires</i> .....</b>	<b>171</b>
<b>Contexte .....</b>	<b>171</b>
<b>Contenu .....</b>	<b>172</b>
<b>Téléchargement .....</b>	<b>173</b>
<b><i>Annexe 2 - Base de données des Effets Sonores de la BBC</i> .....</b>	<b>174</b>
<b>À propos des effets sonores de la BBC .....</b>	<b>174</b>
<b>La collection de l'Unité d'histoire naturelle .....</b>	<b>174</b>
<b><i>Annexe 3 – Base de données des sons de battements cardiaques</i> .....</b>	<b>176</b>
<b>Qu'y a-t-il dans un battement de cœur ? .....</b>	<b>176</b>
<b>Contenu .....</b>	<b>176</b>
<b>Défi d'apprentissage automatique de Peter Bentley et al. ....</b>	<b>176</b>
Contexte .....	177
Aperçu de la tâche .....	177
Téléchargements .....	178
Description et organisation des données .....	179



## Liste des figures

---

<i>Fig. 1.1 - Techniques de dissimulation de l'information</i> .....	2
<i>Fig. 1.2 - Schéma fonctionnel de la cryptographie</i> .....	3
<i>Fig. 1.3 - Dissimulation d'un message secret dans un schéma stéganographique</i> .....	5
<i>Fig. 1.4 - Les phases du cycle de vie du tatouage numérique</i> .....	7
<i>Fig. 2.1 - Données multimédia</i> .....	14
<i>Fig. 2.2 - La forme d'une onde sinusoïdale ou vibration sinusoïdale</i> .....	16
<i>Fig. 2.3 - Les types de fréquence</i> .....	17
<i>Fig. 2.4 - Amplitude en fonction du temps d'un signal apériodique</i> .....	17
<i>Fig. 2.5 - Analyse d'un son complexe</i> .....	18
<i>Fig. 2.6 - Spectre fréquentiel d'un son pur et d'un son complexe périodique</i> .....	18
<i>Fig. 2.7 - Echantillonnage du son</i> .....	21
<i>Fig. 2.8 - Valeurs binaires des intervalles de quantification (quantificateur 3 bits)</i> .....	22
<i>Fig. 2.9 - Schéma de principe d'un Codec audio perceptuel</i> .....	32
<i>Fig. 2.10 - Seuil de silence et seuil de masquage. [43]</i> .....	34
<i>Fig. 2.11 - Caractéristiques du masquage SMR et MNR</i> .....	34
<i>Fig. 2.12 - Illustration des effets du masquage temporel [42]</i> .....	35
<i>Fig. 2.13 - Schéma de principe des codeurs basés sur la perception (Structure of MPEG-1 audio encoder and decoder -Layers I et II)</i> .....	36
<i>Fig. 3.1 - Schéma général d'un tatouage audio numérique</i> .....	40
<i>Fig. 3.2 - Processus d'insertion d'une marque</i> .....	41
<i>Fig. 3.3 - Processus de détection/extraction d'une marque</i> .....	42
<i>Fig. 3.4 - Compromis entre robustesse, capacité et imperceptibilité</i> .....	44
<i>Fig. 3.5 - Classification de tatouage audio numérique</i> .....	45
<i>Fig. 3.6 - Transformation du domaine de Cepstrum</i> .....	47
<i>Fig. 3.7 - Transformation inverse du domaine de Cepstrum</i> .....	47
<i>Fig. 3.8 - Domaines d'algorithmes de tatouage numérique</i> .....	52
<i>Fig. 3.9 - Exemple de technique de base du bit le moins significatif (LSB)</i> .....	53
<i>Fig. 3.10 - Comparaison des distributions non tatouées et tatouées de la différence de moyenne</i> .	57
<i>Fig. 3.11 - Procédure générique d'insertion de la QIM</i> .....	59
<i>Fig. 3.12 - Insertion et extraction de la marque dans le domaine de transformation</i> .....	61
<i>Fig. 3.13 - Distribution de fréquences dans le module de la transformée de Fourier</i> .....	62
<i>Fig. 3.14 - Insertion de la marque dans un domaine DCT basé trame</i> .....	65
<i>Fig. 3.15 - Décomposition DWT à un seul niveau</i> .....	68
<i>Fig. 3.16 - Transformée en ondelettes à trois niveaux</i> .....	68
<i>Fig. 3.17 - Insertion et extraction de la marque dans un domaine de transformation DWT</i> .....	69
<i>Fig. 3.18 - Schéma d'insertion de Devi et Puzgzhentib [130]</i> .....	76
<i>Fig. 3.19 - Schéma d'insertion d'Irawati et al. [138]</i> .....	77
<i>Fig. 3.20 - Schéma d'insertion de Jiang et al. [141]</i> .....	78
<i>Fig. 4.1 - Diagramme de l'oreille (Source Wikipédia)</i> .....	82
<i>Fig. 4.2 - Largeur de bande critique et échelles perceptives résultantes, telles que modélisées par l'échelle de Bark et la largeur de bande rectangulaire équivalente (ERB) [45]</i> .....	83

---

<i>Fig. 4.3 - Le seuil absolu d'audition.</i> .....	84
<i>Fig. 4.4 - Largeur de bande critique à chaque fréquence centrale.</i> .....	85
<i>Fig. 4.5 - Illustrations des effets de masquage spectrale.</i> .....	86
<i>Fig. 4.6 - Exemple de bruit masquant tonalité.</i> .....	87
<i>Fig. 4.7 - Exemple de tonalité masquant bruit.</i> .....	87
<i>Fig. 4.8 - Propriété de masquage temporel du HAS.</i> .....	88
<i>Fig. 4.9 - Schéma fonctionnel générique d'un système de tatouage audio numérique</i> .....	89
<i>Fig. 4.10 - Catégorisation des attaques existantes du tatouage audio [68]</i> .....	91
<i>Fig. 4.11 - Transmission radio d'un signal CD joué</i> .....	94
<i>Fig. 4.12 - Schéma fonctionnel de PEAQ</i> .....	108
<i>Fig. 4.13 - Schéma fonctionnel heuristique de détermination de <math>\alpha</math> dans un contrôle d'imperceptibilité</i> .....	109
<i>Fig. 4.14 - Schéma fonctionnel analytique de détermination de <math>\alpha</math> dans un contrôle d'imperceptibilité</i> .....	109
<i>Fig. 5.1 - Le processus de tatouage</i> .....	122
<i>Fig. 5.2 - Représentation de l'amplitude des coefficients DFT tatoués</i> .....	123
<i>Fig. 5.3 - Représentation de l'amplitude des coefficients FrFT tatoués</i> .....	124
<i>Fig. 5.4 - Représentation de l'amplitude des coefficients QDFT tatoués</i> .....	124
<i>Fig. 5.5 - Schéma DWT à un niveau</i> .....	131
<i>Fig. 5.6 - Différence entre le fichier audio original et le fichier tatoué en LSB</i> .....	133
<i>Fig. 5.7 - Différence entre le fichier audio original et le fichier tatoué en DCT</i> .....	134
<i>Fig. 5.8 - Différence entre le fichier audio original et le fichier tatoué en DWT</i> .....	135
<i>Fig. 5.9 - SNR obtenu pour les trois domaines d'insertion</i> .....	136
<i>Fig. 5.10 - Différence entre le fichier audio original et le fichier tatoué à l'aide de Sch2.V2</i> .....	142
<i>Fig. 5.11 - Différence entre le fichier audio original et le fichier tatoué à l'aide de Sch2.V2</i> .....	143
<i>Fig. 5.12 - Différence entre le fichier audio original et le fichier tatoué à l'aide de Sch1.V2</i> .....	144

## Liste des tableaux

---

Tableau 2.1 - La relation entre les décibels, les rapports d'intensité et les rapports de pression .	20
Tableau 2.2 - Taux d'échantillonnage de certaines applications audio .....	21
Tableau 2.3 - Formats audio non compressés .....	25
Tableau 2.4 – Formats audio compressés sans perte .....	26
Tableau 2.5 – Formats audio compressés avec perte .....	27
Tableau 2.6 - Paramètres de base des classes du signal audio acoustique .....	29
Tableau 2.7 - Débits binaires CD et DAT.....	30
Tableau 3.1 - Comparaison entre domaine spatial et domaine fréquentiel.....	46
Tableau 3.2 - Types de tatouage audio basés sur différents critères .....	51
Tableau 3.3 - Récapitulatif des caractéristiques des techniques citées .....	74
Tableau 4.1 - Répartition fréquentielle des 24 bandes critiques de l'oreille humaine .....	84
Tableau 4.2 - Résumé et abréviations des attaques existantes .....	90
Tableau 4.3 - Descriptions des tests de robustesse de base.....	97
Tableau 4.4 - Score d'opinion moyen.....	104
Tableau 4.5 - Valeurs expérimentales d'une mesure MOS subjective .....	105
Tableau 4.6 - Echelle de dégradation à cinq notes et valeur de la SDG associée.....	106
Tableau 4.7 - Echelle de dégradation à cinq notes et valeur de l'ODG associée .....	107
Tableau 4.8 - Niveaux d'assurance possibles pour l'imperceptibilité.....	112
Tableau 5.1 - Règles de substitution pour la première variante .....	122
Tableau 5.2 - Règles de substitution pour la deuxième variante .....	122
Tableau 5.3 - Résultats d'imperceptibilité.....	125
Tableau 5.4 - Comparaison des résultats d'imperceptibilité avec des travaux connexes .....	125
Tableau 5.5 - Comparaison de la robustesse avec des travaux connexes .....	126
Tableau 5.6 - Condition de redistribution des valeurs de l'échantillon .....	129
Tableau 5.7 - Condition de redistribution des coefficients DCT .....	130
Tableau 5.8 - Condition de redistribution des coefficients AC.....	131
Tableau 5.9 - Résultats d'imperceptibilité pour l'insertion dans le domaine spatial.....	133
Tableau 5.10 - Résultats d'imperceptibilité pour l'insertion dans le domaine de la fréquence .....	134
Tableau 5.11 - Résultats d'imperceptibilité pour l'insertion dans le domaine multi-résolution ...	135
Tableau 5.12 - Comparaison de l'imperceptibilité avec des travaux connexes .....	136
Tableau 5.13 - Comparaison de la robustesse avec les travaux connexes .....	137
Tableau 5.14 - Premier schéma d'insertion proposé .....	139
Tableau 5.15 - Deuxième schéma d'insertion proposé .....	139
Tableau 5.16 - Résultats d'imperceptibilité pour l'insertion dans le domaine spatial.....	143
Tableau 5.17 - Résultats d'imperceptibilité pour l'insertion dans le domaine fréquentiel .....	144
Tableau 5.18 - Résultats d'imperceptibilité pour l'insertion dans le domaine multi-résolution ....	145
Tableau 5.19 - Comparaison de l'imperceptibilité avec des travaux connexes .....	145
Tableau 5.20 - Comparaison de la robustesse avec des travaux connexes .....	146

# 1. Introduction Générale

---

## Sommaire

---

1.1	Les techniques de dissimulation de l'information .....	2
1.2	Un aperçu du tatouage numérique .....	6
1.3	Problématique ..	8
1.4	Organisation de la thèse .....	10

---

## 1.1 LES TECHNIQUES DE DISSIMULATION DE L'INFORMATION

La dissimulation d'informations est un concept général qui consiste à cacher des données dans un contenu. Le terme " cacher " peut être interprété comme le fait de garder secrète l'existence de l'information ou de la rendre imperceptible [1].

Le terme dissimulation ou masquage peut faire référence soit à l'information secrète (Stéganographie) soit à l'imperceptibilité de l'information (Tatouage). Deux sous-disciplines importantes de la dissimulation de l'information sont la stéganographie et le tatouage qui sont étroitement liées l'une à l'autre mais avec des propriétés, des exigences et des conceptions différentes.

La stéganographie cherche des moyens de rendre la communication invisible ou secrète en cachant des secrets dans un message de couverture, tandis que le tatouage trouve son origine dans la nécessité de protéger le contenu par des droits d'auteur [2]. Les techniques générales de la dissimulation de l'information sont illustrées dans la figure 1.1 [3]. Ces techniques sont utilisées pour sécuriser les documents numériques transmis sur les réseaux, la cryptographie est une discipline pour sécuriser les données, mais diffère des autres techniques de dissimulation de l'information à savoir la Stéganographie et le tatouage. Dans les sections suivantes nous présentons brièvement ces techniques en montrant les différences qui existent entre leurs méthodes de dissimulation utilisées.

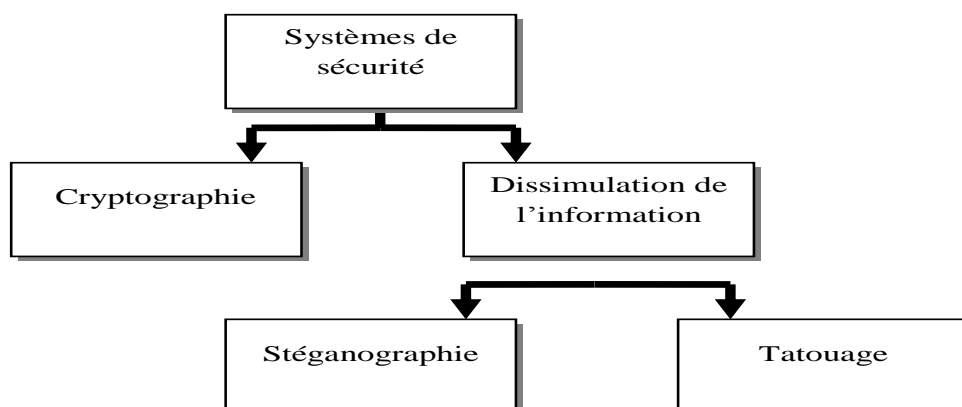


Fig. 1.1 - Techniques de dissimulation de l'information

### 1.1.1 La Cryptographie

Le terme cryptographie signifie écriture secrète ou cachée. Il s'agit de l'étude des méthodes de transmission des messages sous différentes formes afin que seuls les destinataires visés puissent enlever le chiffrement et lire le message. Le message qu'on veut le rendre secret est appelé texte clair, et le message chiffré ou caché est appelé texte chiffré. Le chiffrement ou le cryptage est le processus de conversion du texte clair en texte chiffré et le déchiffrement ou le décryptage est le processus inverse qui convertit un texte chiffré inintelligible en texte clair. Le schéma fonctionnel de la cryptographie est présenté dans la figure 1.2 [3].



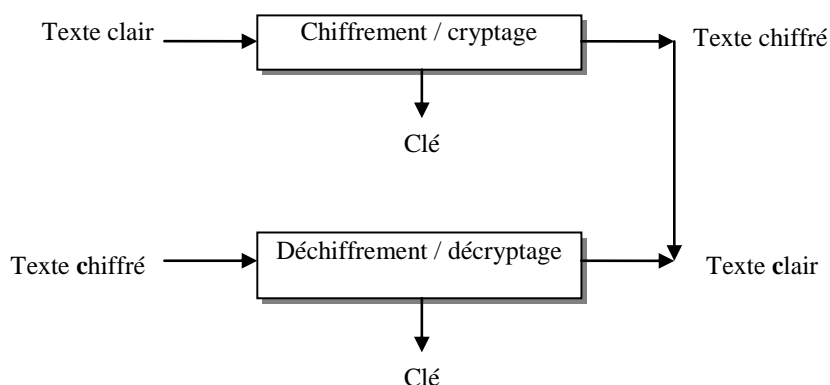


Fig. 1.2 - Schéma fonctionnel de la cryptographie

Dans [4] on trouve les définitions suivantes:

- **Chiffrement ou cryptage** : Le processus de verrouillage des informations à l'aide de la cryptographie est le cryptage. L'information verrouillée est un message crypté.
- **Clé** : Un secret, similaire à un mot de passe, qui est utilisé pendant le chiffrement et le déchiffrement.
- **Déchiffrement ou décryptage** : Le processus de déverrouillage des informations cryptées à l'aide de la cryptographie.
- **Un crypto système** : est une mise en œuvre de techniques cryptographiques et de l'infrastructure qui les accompagne pour fournir des services de sécurité de l'information. Deux sciences sont liées aux crypto systèmes :
  1. **Cryptographie** : C'est l'art de rendre le cryptosystème capable d'assurer la sécurité de l'information.
  2. **Cryptanalyse** : C'est l'art de casser le texte chiffré. La cryptographie traite le texte chiffré pour la transmission ou le stockage. La cryptanalyse implique l'étude des mécanismes cryptographiques dans le but de les casser. Elle est également utilisée lors de la conception de nouvelles techniques cryptographiques pour tester leurs forces de sécurité.

Un crypto système est dit symétrique si la clé utilisée pour le déchiffrement est la même que celle utilisée pour le chiffrement. Il est asymétrique s'il utilise une paire de clés, une est publique connue par tout le monde et utilisée pour le chiffrement et la vérification de la signature numérique, l'autre est privée, et est utilisée pour le déchiffrement et la génération de la signature numérique. Dans une communication réseau, des crypto systèmes hybrides sont utilisés, un crypto système asymétrique pour générer une clé de session et un crypto système symétrique utilise cette clé pour chiffrer la communication de bout en bout.

De nombreuses méthodes ont été développées pour chiffrer et déchiffrer les messages afin de les garder confidentiels. La cryptographie garde le contenu du message confidentiel durant sa transmission seulement, mais elle est incapable de garder l'existence du message en secret. La procédure de chiffrement vise principalement à protéger le message texte (ou tout autre type de données) lors de sa transmission, une fois déchiffré, le message n'est plus protégé.

---

### 1.1.2 La Stéganographie

Le mot stéganographie est dérivé du grec "*steganos*" et "*graphia*", qui signifient littéralement "écriture couverte". Comme défini dans [2], la stéganographie fait référence à la pratique consistant à modifier de manière indétectable une couverture pour y intégrer un message secret, c'est-à-dire à transmettre des informations cachées de telle sorte que personne, hormis l'expéditeur et le destinataire, ne soupçonne l'existence même du message. La stéganographie a été utilisée de nombreuses manières à travers le temps, par exemple, les tatouages cachés, les encres invisibles, les micropoints, la disposition des caractères, les chiffrements nuls, les mots de code, les canaux secrets et la communication à spectre étalé [5].

La stéganographie est l'art de la science de la communication invisible. Elle cache l'information secrète dans d'autres informations de telle manière que l'information cachée ne semble rien à l'œil humain. En plus d'établir une communication cachée entre un émetteur et un receveur via un signal audio ou une image, la stéganographie cherche à cacher l'existence de la communication à une tierce.

Il faut noter que la stéganographie semble être comparable à la cryptographie, mais pas synonyme. La cryptographie et la stéganographie sont toutes deux des moyens d'assurer le secret, mais leurs méthodes de dissimulation sont différentes. En cryptographie, le message est chiffré pour protéger son contenu. On peut dire qu'un message a été chiffré, mais on ne peut pas le déchiffrer sans le bon code. Une fois que les données sont déchiffrées, la protection est supprimée et il n'y a plus de confidentialité. En stéganographie, le message existe, mais sa présence est inconnue du récepteur et d'autres personnes, comme l'attaquant ou le pirate. C'est grâce à ce manque d'attention que le secret est bien préservé. Comme indiqué dans [2], "*Un message cryptographique peut être intercepté par une personne écouteur (eavesdropper), cependant, cette personne ne peut même pas savoir qu'un message stéganographique existe*". Par conséquent, la stéganographie ne protège pas seulement les informations confidentielles, comme le fait la cryptographie, mais permet également de protéger les parties communicantes dans une certaine mesure. En plus, la stéganographie et la cryptographie peuvent être combinées pour fournir deux niveaux de sécurité. C'est-à-dire nous pouvons chiffrer un message à l'aide de la cryptographie, puis nous cachons le chiffrement dans la couverture en utilisant la stéganographie. Cette notion peut être adoptée dans le système de tatouage numérique pour accroître la sécurité. La figure 1.3 illustre la dissimulation d'un message qui peut être chiffré à l'aide d'une clé dans un support de couverture (image, vidéo, audio,...), le message peut être chiffré à l'aide d'une clé utilisée dans les processus d'encodage et du décodage.

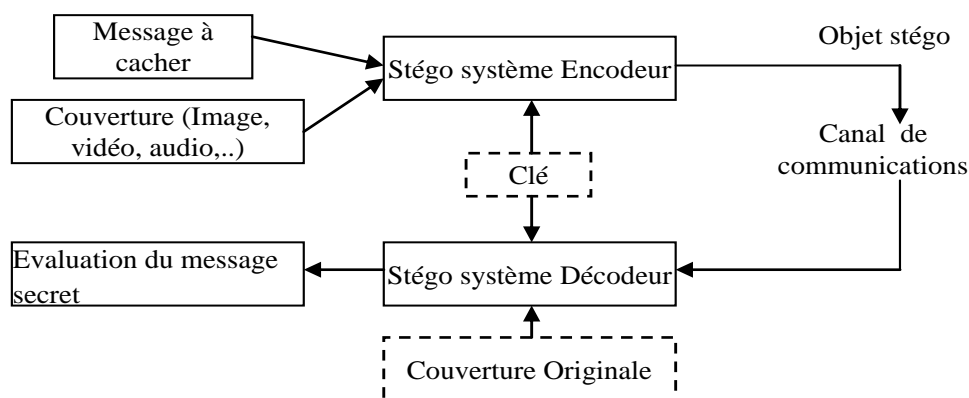


Fig. 1.3 - Dissimulation d'un message secret dans un schéma stéganographique

Dans la figure 1.3 Le message secret et le message de couverture sont d'abord transmis au codeur à l'aide de techniques de stéganographie, un chiffrement éventuel peut être utilisé et dans ce cas le même algorithme de chiffrement est mis en œuvre dans l'ordre inverse pour récupérer le message secret. La clé joue un rôle très important dans la récupération du message original.

La stéganographie peut être divisée en deux types : Fragile et Robuste [6]

- **Fragile** : Elle consiste à insérer des informations dans un fichier qui seront détruites si le fichier est modifié. Cette méthode ne convient pas pour enregistrer le détenteur du droit d'auteur du fichier, car elle peut être si facilement supprimée. Elle est utile dans les situations où il est important de prouver que le fichier n'a pas été altéré. Cette technique est considérée comme plus facile à mettre en œuvre que les méthodes robustes.
- **Robuste** : Les informations insérées ne peuvent pas être facilement détruites. La marque doit être cachée dans une partie du fichier où sa suppression serait facilement perçue.

### 1.1.3 Le Tatouage

Le tatouage est une pratique qui consiste à modifier imperceptiblement un objet pour y intégrer un message [1], c'est-à-dire à cacher des informations spécifiques sur l'objet sans distorsion perceptible. Le tatouage a une longue histoire qui remonte à la fin du treizième siècle, lorsque les papeteries italiennes ont inventé les "*watermarks*" pour indiquer la marque ou le fabricant du papier et ont également servi de base à l'authentification du papier. Au XVIIIe siècle, les "*watermarks*" ont commencé à être utilisés comme mesures anti-contrefaçon sur les billets de banque et autres documents. À ce jour, la forme la plus courante de "*watermark*" sur papier reste le billet de banque dans de nombreux pays. Le premier exemple d'une technologie similaire à notre concept de tatouage - des informations imperceptibles sur les objets dans lesquels elles sont incorporées, C'est ce qu'a fait *Emil Hembrooke* en 1954 en tatouant des œuvres musicales. Il a inséré un code Morse pour identifier la propriété de la musique, de sorte que toute infraction puisse être discernée. Le terme "*tatouage numérique*" est le résultat de l'ère numérique, qui semble avoir été utilisé pour la première fois par *Komatsu et Tominaga* en 1988. Mais c'est en 1993 que Tirkel a présenté deux techniques de tatouage pour cacher les données dans les images [7] et a utilisé le terme "tatouage numérique" (Digital Watermarking en anglais). A partir de

cette date, le tatouage numérique a suscité beaucoup d'intérêt et a évolué très rapidement [1], [2].

Le tatouage numérique peut être défini comme le processus d'insertion d'une certaine quantité d'informations (techniquement appelé marque ou filigrane) dans le contenu multimédia, y compris les documents texte, les images, les flux audio ou vidéo, de telle sorte que la marque puisse être détectée ou extraite ultérieurement pour faire une assertion ou une preuve sur les données. Il permet à une personne de fournir des avis cachés de droit d'auteurs ou d'autres moyens de vérification. Son objectif est de protéger le document numérique contre le piratage, par l'insertion d'une marque ou une signature. Il régit la propriété, l'authenticité et les droits d'utilisation du document. Le tatouage vise à adapter la marque en fonction de l'action de ce pirate sur le document : la marque doit être ineffaçable si l'on souhaite lutter contre le piratage des documents, à l'inverse, elle doit disparaître dès que l'attaquant modifie le document si l'on souhaite garantir l'intégrité du document.

Le tatouage et la stéganographie sont deux domaines de la dissimulation de l'information avec des aspects différents. Tous les deux doivent être robustes pour protéger le message secret. Cependant, le secret dans le tatouage n'est pas strictement nécessaire, alors que la stéganographie doit être secrète par définition. Par exemple, il est préférable que tout le monde connaisse la présence d'une marque sur les billets de banque et puisse la reconnaître facilement à la lumière. La stéganographie ne requiert qu'une robustesse limitée, car elle concerne généralement la communication secrète bidirectionnelle entre des parties de confiance. Tandis que le système de tatouage doit être assez robuste pour résister à toute tentative de suppression des données secrètes, car les objets tatoués sont ouverts au public. De plus, le message caché dans le tatouage est lié à l'objet, qui a la même importance que l'objet lui-même. Par conséquent, aucune détérioration de la qualité perceptive de l'objet n'est souhaitée. Mais en stéganographie, l'objet n'est qu'un support pour véhiculer le message et n'a aucune valeur intrinsèque [1], [2].

## 1.2 UN APERÇU DU TATOUAGE NUMERIQUE

Le tatouage numérique est un domaine scientifique récent apparu au début des années 90 et depuis il a connu un essor spectaculaire. Le terme « tatouage numérique » (Digital Watermarking en anglais) est apparu pour la première fois en 1993, lorsque Tirkel a présenté deux techniques de tatouage pour cacher les données dans les images [7]. Durant les deux dernières décennies, la distribution en ligne de multimédia numérique comprenant des images, audio, vidéos et documents a augmenté rapidement. Dans l'environnement ouvert de l'internet, il est facile d'obtenir un accès gratuit à diverses ressources d'information. Outre la commodité et la haute fidélité avec lesquelles les données numériques peuvent être copiées, modifiées et transmises, un grand nombre d'infractions aux droits d'auteur sont apparues à la suite de la reproduction et de la redistribution non autorisées, ce qui empêche l'industrie du multimédia numérique de progresser régulièrement [8]. Pour prévenir ces violations, l'application de la gestion de la propriété est devenue une nécessité urgente et suscite de plus en plus d'attention. En conséquence, le tatouage numérique a été proposé pour identifier le propriétaire ou le distributeur de données numériques dans le but de protéger les droits d'auteur. Initialement le tatouage

numérique est venu pour renforcer la protection des droits d'auteur du contenu multimédia, mais il tend de plus en plus à remplir d'autres fonctions de sécurité.

La numérisation de tous les domaines technologiques a considérablement élargi la notion de tatouage, et de nombreuses nouvelles possibilités ont été ouvertes. En particulier, il est possible de cacher des informations dans des fichiers numériques d'images, d'audio et de vidéo d'une manière non perçues et statistiquement indétectables. Motivée par les préoccupations relatives à la gestion des droits numérique (*digital rights management*), une nouvelle technique appelée "tatouage numérique" a été proposée pour protéger la propriété intellectuelle et le droit d'auteur [9, p. 3]. Le tatouage numérique n'est pas conçu pour révéler la relation exacte entre le contenu protégé par le droit d'auteur et les utilisateurs, à moins que l'un d'eux ne viole son utilisation légale.

Le tatouage numérique est le processus d'insertion imperceptible de marque(s) dans des médias numériques comme des signes permanents, puis d'extraire la(s) marque(s) du média suspect pour en assurer l'authenticité [9, p. 3]. La(s) marque(s) est (sont) toujours associée au média numérique à protéger ou à son propriétaire, ce qui signifie que chaque support numérique a sa propre marque ou que chaque propriétaire a sa seule marque. Pour la protection des droits d'auteur, l'avantage du tatouage numérique par rapport à la stéganographie et à la cryptographie traditionnelles est que les médias numériques peuvent être utilisés ouvertement, malgré la présence de marques. En d'autres termes, nous ne limitons pas l'accès aux marques résidant dans les supports numériques, mais nous faisons des efforts supplémentaires pour améliorer leur robustesse contre diverses attaques.

Le cycle de vie d'un tatouage numérique est illustré à la figure 1.4.

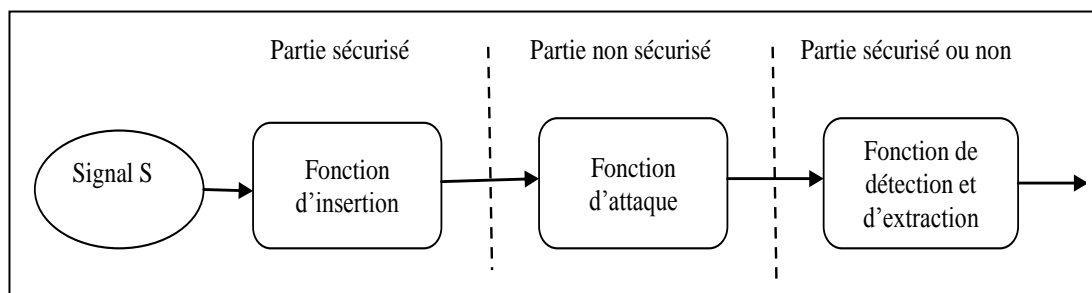


Fig. 1.4 - Les phases du cycle de vie du tatouage numérique

Durant ce cycle de vie, les données sont insérées avec le "watermark" et peuvent circuler à travers le canal (qui n'est pas sûr). Et finalement, la marque est extraite pour vérifier l'authenticité des données [7]. Le tatouage est de deux types : le tatouage visible et le tatouage invisible.

- **Tatouage visible** : Il s'agit de l'information visible sur les données. Les marques visibles sont généralement des logos ou du texte.
- **Tatouage invisible** : Il n'est pas visible ou perceptible, mais il peut être détecté par différents moyens. Il peut également s'agir d'une forme ou d'un type de stéganographie et est utilisé à grande échelle. Il peut être récupéré facilement.

Il convient de mentionner que certains chercheurs fournissent un autre terme étroitement lié à la question de la protection des droits d'auteur, ce qu'on appelle les empreintes digitales numériques ou les signatures (digital fingerprinting) [10]. Les empreintes digitales sont des caractéristiques d'un objet qui tendent à le distinguer d'autres objets similaires. Au sens strict, les empreintes digitales désignent le processus d'identification et d'enregistrement des empreintes digitales qui sont déjà intrinsèques à l'objet [2]. Elle est souvent considérée comme une forme de tatouage légal utilisé pour retracer les utilisateurs autorisés et qui distribuent illicitement l'objet, c'est-à-dire le problème de traçage des copies distribuées illégalement. Il faut noter les plus grandes différences entre le tatouage numérique et la signature numérique sont l'origine des messages cachés et le mode de fonctionnement. Dans le tatouage numérique, la marque est un message arbitraire contenant l'information sur la propriété, tandis que l'empreinte digitale dans la signature numérique est dérivée de l'hôte lui-même et produit un métafichier qui décrit le contenu du fichier source, de sorte qu'une partie de travail peut être facilement trouvée et comparée à d'autres travaux dans la base de données [11]. Pour cette raison, les signatures numériques ont d'abord été conçues pour la recherche rapide. D'une manière un peu différente, le tatouage numérique est issu de la motivation pour la protection des droits d'auteur des produits multimédias numériques. Il est capable d'être un outil efficace pour l'application de ces droits d'auteur.

### 1.3 PROBLEMATIQUE

La croissance et le succès récent de l'Internet, ainsi que la disponibilité de dispositifs d'enregistrement et de stockage numériques relativement accessibles et peu coûteux [12], ont créé un environnement dans lequel il devient très facile d'obtenir, de reproduire et de distribuer du contenu numérique sans perte de qualité. Cette situation est devenue une préoccupation majeure pour les industries d'édition de contenu multimédia (son, vidéo et image), car il n'existait pas de technologies ou de techniques permettant de protéger les droits de propriété intellectuelle des médias numériques et d'empêcher leur duplication non autorisée [13]. Le transfert de données sensibles ne peut se faire avec de tels risques et doit donc être sécurisé. Plusieurs méthodes ont été déployées pour pallier à ce problème. La cryptographie a été l'outil efficace utilisé pour protéger et sécuriser les informations en transmettant des données chiffrées. Elle offre ainsi des solutions pour protéger la confidentialité des données, pour assurer leur intégrité ou pour connaître l'identité de la personne qui les envoie. Mais ces protections n'agissent que pendant la transmission et la diffusion du média numérique. Une fois que les données sont en clair, elles ne contiennent plus aucune protection. Les technologies de chiffrement peuvent être utilisées pour empêcher l'accès non autorisé au contenu numérique. Cependant, le chiffrement a ses limites en matière de protection des droits de propriété intellectuelle, car une fois le contenu numérique déchiffré, rien n'empêche un utilisateur non autorisé de le reproduire illégalement [14]. Pour surmonter ce problème, des techniques de dissimulation de données telle que la stéganographie permet de dissimuler des informations dans un fichier et de maintenir ainsi le message secret protégé même après réception du fichier. La stéganographie est une communication secrète bidirectionnelle, elle est généralement appliquée à une relation un à un. Cependant l'objectif de la steganographie est de protéger le message secret à transmettre et non le fichier média numérique utilisé comme couverture pour le message.

Une autre technologie moderne est manifestement nécessaire pour aider à la fois à établir et à prouver simplement les droits de propriété, à surveiller de manière adéquate l'utilisation possible du contenu, à garantir l'accès autorisé, à faciliter l'authentification du contenu et à empêcher typiquement la reproduction illégale, et surtout avec la portée considérable et le grand potentiel d'informations spécifiques dans le domaine de multimédia numérique dans de nombreux aspects pratiques de la vie quotidienne et professionnelle [15], la télévision, l'Internet, l'audiovisuel, l'imagerie médicale ou satellitaire et la surveillance à distance. Ce besoin spécifique a attiré l'attention de la communauté des chercheurs et de l'industrie, ce qui a conduit à la création d'une nouvelle forme de dissimulation de l'information, appelée tatouage numérique [16], qui reste la méthode la plus pratique pour résoudre le problème de la confidentialité et du droit d'auteur et garantir l'authenticité des produits multimédias. L'idée fondamentale du tatouage numérique est de créer des métadonnées contenant des informations précises sur le contenu numérique à protéger [17], puis de cacher soigneusement les métadonnées dans ce contenu. Les informations stockées sous forme de métadonnées peuvent utiliser différents formats, comme un pseudo bruit gaussien, une chaîne de caractères ou une image binaire, voire une image numérique.

Dans ce domaine du tatouage numérique, beaucoup de travaux ont été réalisés pour augmenter la capacité des données à dissimuler tout en préservant une imperceptibilité raisonnable et ceux dans les différents domaines d'insertion. La rapidité d'exécution de ces algorithmes les rend attractives pour les transmissions sécurisées en temps réel sur les réseaux de communication.

Un système de tatouage efficace doit garantir l'imperceptibilité des données cachées [18]. La marque insérée ne doit pas affecter la qualité perceptive de la couverture. L'imperceptibilité, la robustesse et la capacité de la marque sont les propriétés importantes qui caractérisent la technique de tatouage. La robustesse garantit que la marque imperceptible ne peut être détruite sans que le support ne soit gravement dégradé, ce critère représente la résistance après des changements soudains par des attaques [19]. La transparence ou l'imperceptibilité indique que, même lors de l'insertion d'informations secrètes, le document tatoué n'est perceptivement pas perturbé. La capacité est également une mesure importante pour évaluer l'efficacité d'un tatouage. Elle indique la quantité d'informations pouvant être insérée sans être audible et sans que le support de couverture ne soit dégradé [20]. Ces propriétés sont étroitement liées mais ont l'effet inverse. Si la capacité est grande, elle a un effet positif sur la robustesse mais un effet négatif sur l'imperceptibilité et vice versa. Il faut nécessairement trouver un compromis entre ces trois paramètres en fonction de l'application envisagée.

Afin de sécuriser les données échangées en télémédecine, le but de notre travail est de proposer des approches irréversibles de tatouage audio numérique aveugle, qui seront robustes et imperceptibles. Le schéma de l'extraction de la marque s'effectue sans l'utilisation du fichier original ni de la marque. Ces approches seront appliquées aux fichiers audio contenant des informations médicales comme les battements de cœurs et les sons respiratoires. Ce tatouage doit être réalisé par la dissimulation d'informations secrètes dans le fichier hôte et la marque insérée par un tel schéma de tatouage doit être ineffaçable et robuste face aux différentes attaques intentionnelles ou non intentionnelles, il sera impossible d'enlever la marque insérée sans avoir recours à la clé ou à la méthode d'insertion. La marque insérée est composée des

informations du patient et des données d'acquisition. Ainsi, lors du processus d'extraction, une comparaison entre les informations du patient et la marque extraite indiquera si les données ont été modifiées. Cela permettra de s'assurer que l'échantillon audio correspond bien au patient indiqué dans l'enregistrement.

Chaque approche proposée a été déclinée en plusieurs variantes qui ont été analysées et comparées selon la capacité de dissimulation pour chaque variante et selon l'imperceptibilité en analysant pour chaque méthode les mesures quantitatives de distorsion obtenues. Les expériences ont été réalisées sur la robustesse dans l'insertion de la marque et la résistance aux attaques comme l'ajout du bruit, le ré-échantillonnage, le filtrage et la compression. Enfin nos approches ont été comparées à certains travaux connexes récents sur le plan d'imperceptibilité, robustesse et capacité d'insertion.

## 1.4 ORGANISATION DE LA THESE

Le travail présenté dans cette thèse est divisé en six chapitres. Le premier chapitre tient lieu d'introduction générale au reste du document. Le deuxième chapitre est dédié au son et le traitement numérique des fichiers audio. Les deux chapitres suivants sont dédiés à l'état de l'art qui concerne les techniques de tatouage audio numérique et les méthodes d'évaluation de ces schémas de tatouage existant dans la littérature, cette étude est importante pour la bonne compréhension du contexte de travail. Le cinquième chapitre est consacré à la présentation de nos travaux de contributions qui ont été réalisés dans le cadre de cette thèse. Nos conclusions générales sont présentées au dernier chapitre.

Le chapitre 1 permet notamment de présenter le contexte de cette thèse, de situer la problématique abordée et montrer l'objectif des travaux réalisés.

Le chapitre 2 permet de présenter le son, ses caractéristiques physiques, sa numérisation, sa compression en tant que signal audio et les différents formats des fichiers audio non compressés et compressés sans perte et avec perte. Il présente également certaines des idées de base des concepts du codage audio perceptuel qui permettent de réaliser des réductions importantes de débit binaire dans la représentation de matériel audio de haute qualité.

Le chapitre 3 présente le contexte technique qui entoure le tatouage audio numérique et les différents travaux dans la littérature qui ont fait avancer la technologie dans ce domaine. Il présente l'architecture générale du système de tatouage audio numérique avec ses propriétés et ses conditions requises, il dresse une taxonomie des techniques de tatouage audio classé selon différents critères : le domaine d'insertion, la perception humaine, la technique d'insertion, le champ d'application et la méthode de détection et d'extraction. En plus, ce chapitre présente les différentes techniques de tatouage audio dans les différents domaines d'insertion avec une analyse comparative entre ces techniques et en montrant leurs avantages et leurs inconvénients. En fin, il passe en revue les différents travaux récents que nous avons utilisés pour la comparaison de nos approches réalisées dans le cadre de cette thèse.



Le chapitre 4 présente les propriétés du système auditif humain à prendre en considération dans le codage et le tatouage audio pour garantir l'imperceptibilité des schémas de tatouage. Il montre également des classifications importantes des attaques audio « intentionnelles et non intentionnelles » pouvant être réalisées sur les systèmes de tatouage audio numérique ainsi que les outils et les tests cités dans la littérature permettant d'évaluer ces systèmes face à ces attaques. Enfin, il présente différents aspects d'évaluation des systèmes de tatouage audio numérique.

Le chapitre 5 présente nos approches de tatouage numérique aveugle pour protéger des fichiers audio du domaine de la télémédecine et qui ont été réalisées dans le cadre de cette thèse. Ces approches ont utilisé des techniques de tatouage du domaine spatial, fréquentiel et du domaine de multi-résolution. Chaque approche a été déclinée en plusieurs variantes qui ont été analysées et comparées selon la capacité de dissimulation pour chaque variante et selon l'imperceptibilité et la robustesse en analysant pour chaque méthode les mesures quantitatives de distorsion obtenues.

Nous concluons dans le chapitre 6 par un rappel des différentes contributions de notre thèse, Nous montrons à la fin des perspectives pour des travaux futurs permettant de poursuivre les recherches réalisées dans cette thèse.

## 2. Le Son et le Traitement Numérique des Fichiers Audio

---

### Sommaire

---

2.1	<b>Introduction</b>	13
2.2	<b>Le son numérique</b>	15
2.3	<b>Format de fichier audio</b>	24
2.4	<b>Codage audio perceptif utilisant des modèles psycho-acoustiques</b>	28
2.5	<b>Conclusion</b>	36

---

## 2.1 INTRODUCTION

D'un point de vue physique, un son est une énergie qui se propage sous forme de vibrations dans un milieu compressible (dans l'eau, dans l'air, dans les matériaux solides).

Le traitement du son est la branche du traitement du signal [21] qui s'applique aux signaux audio, dans le but notamment d'améliorer la qualité, de les compresser, ou d'extraire de l'information. Le traitement du signal développe et étudie les techniques de traitement, d'analyse et d'interprétation des signaux. Plusieurs types d'opérations sont possibles sur ces signaux, comme le filtrage, la compression de données, la numérisation, le codage, le chiffrement et la transmission de données. La science qui étudie les sons s'appelle l'acoustique. La psycho acoustique combine l'acoustique avec la physiologie et la psychologie, pour déterminer la manière dont les sons sont perçus par l'oreille humaine et interprétés par le cerveau.

Le son numérique fait parti d'un domaine de création et de diffusion de l'information qu'est le multimédia. Avant d'aborder le son numérique, les fichiers audio et leur format du codage, il est utile de préciser quel sera le champ couvert par les termes "multimédia". Dans la section suivante, nous donnons quelques définitions qui permettent de cerner ce champ.

### 2.1.1 Définitions Média et Multimédia

A l'origine "Multimédia" est un adjectif qui appartient au vocabulaire de la publicité. Une "compagne multimédia" désigne alors le recours complémentaires à plusieurs médias (presse écrite, affiches, radio et télévision) pour assurer la promotion auprès du grand public d'un nouveau produit, d'un événement culturel, d'un événement politique, etc. Multimédia devient nom avec l'avènement du numérique et des applications multimédias intégrant sur un même support plusieurs types de médias (texte, image, graphique, son, vidéo). Multimédia fait alors référence à une pluralité de médias conjointement utilisés pour la réalisation d'un unique produit de communication, d'apprentissage ou de divertissement.

Média est un moyen physique par lequel les données sont perçues, représentées, stockées ou transmises, médias discrets comme le texte, l'image, et médias continues comme le son et la vidéo [22].

Le mot "multimédia" est apparu vers la fin des années 1980, lorsque les supports amovibles de stockage de données de grandes tailles se sont développés comme les CD-ROM. Plus tard il désignera les applications qui — grâce à la capacité de stockage d'information du CD-ROM, ou d'Internet (par téléchargement), et aux capacités de l'ordinateur — peuvent créer, utiliser ou piloter différents médias simultanément : musique, son, image, vidéo, et interface graphique interactive (IHM). Le terme s'est développé par la suite pour désigner une création utilisant plusieurs médias : l'image, le son et la vidéo (film ou clip). Depuis, son sens s'est élargi pour englober les logiciels, matériels et contenus éditoriaux interactifs mettant en œuvre l'image fixe ou animée, le son, le texte et l'hypertexte [23].

On peut dire que MULTIMEDIA est l'élaboration de documents comprenant de MULTiples MEDIAS comme illustré dans la figure 2.1.

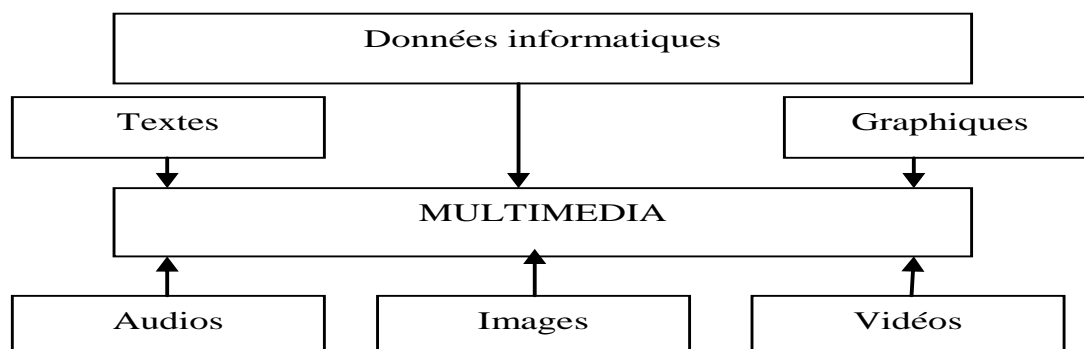


Fig. 2.1 - Données multimédia

### 2.1.2 Médias discrets / Médias continus

On distingue deux types de médias [22]:

- **Médias discrets (ou passifs)** : ils n'ont aucune dépendance temporelle, comme le texte, les graphiques, les images fixes, l'information contenue est donnée exclusivement par l'ensemble des éléments individuels qui le compose, sans aucune référence temporelle. La lecture d'un média discret n'est pas conditionnée par des contraintes temporelles, par conséquent son volume est limité.
- **Médias continus (ou actifs)** : ils ont une dépendance temporelle, comme le son, la musique, la vidéo. Le terme continu se réfère à la perception de l'utilisateur, pas à la représentation interne. L'information qu'ils contiennent ne peut se comprendre que par rapport aux évolutions temporelles de grandeurs physiques. La lecture d'un média continu est critique du point de vue temporel. Par exemple, un groupe d'échantillons audio n'ont pas de sens s'ils sont lus après ceux qui les suivaient initialement. Le volume d'un média continu n'est limité que par la durée, et son débit est en général élevé.

Une caractéristique importante d'un système multimédia est un traitement intégré de médias ayant une dépendance temporelle "*médias continus*" et de médias ayant une indépendance temporelle "*médias discrets*".

### 2.1.3 Caractéristiques des médias continus

Les signaux manipulés dans les systèmes multimédias, et notamment les signaux temporels, ne correspondent pas à des données ordinaires. Elles sont caractérisées par [22]:

- un volume très important. Les débits de la vidéo (non comprimée) se situent entre 140 et 216 Mbits/s, de l'audio type CD de l'ordre de 1.4 Mbits/s, la parole 64 kbits/s. Il est donc nécessaire de compresser ces médias pour réduire les débits de leur lecture.
- un flux continu, ce qui caractérise les médias continus, variable du plus faible (audio) au plus important (vidéo). Il en découle des aspects très importants de temps réel, synchronisation et contrôle des flux.
- des qualités de services spécifiques. Par exemple, en terme de taux d'erreur, si la vidéo tolère de taux modérés, l'audio est plus exigeant et nécessite des taux d'erreur plus faibles.

## 2.1.4 Numérisation

Créer un produit multimédia, c'est : traiter, structurer, synchroniser sur un même support des : Textes, Sons, Images (fixes ou animées), etc. Tout en accordant au futur utilisateur de réelles possibilités d'interaction (chemin, rythme de consultation). Rassembler et synchroniser sur un même support des médias aussi différents discrets et continus, sonores et visuels, fixes et animés. Ce ci implique nécessairement qu'ils partagent le même mode de représentation.

Dans le multimédia, tous les médias (texte, son, image F/A, animations 2D/3D, vidéo...) partagent la même représentation numérique. Cette représentation est susceptible d'être interprétée par le même support/appareil de restitution : Ordinateur, Assistant personnel PDA, Smartphone, Téléphone mobile, Lecteur DVD, etc.

Les techniques de numérisation sont complexes mais le principe général est relativement simple : Il s'agit de transformer une variation continue (sons, niveaux de gris, couleurs, etc.) en une succession d'éléments discontinus (formés de 0 et de 1), dont on pourra représenter ensuite numériquement (les caractéristiques quantifiables). Une représentation numérique binaire de l'objet multimédia exprimée avec un poids en KiloOctet, MegaOctet, GigaOctet, etc...

## 2.2 LE SON NUMERIQUE

### 2.2.1 Les caractéristiques physiques du son

#### 2.2.1.1 La nature du son

Le son provient de la vibration d'un objet. Cette vibration est réalisée sur le milieu environnant (généralement l'air) sous la forme de changements de pression. Les particules de l'atmosphère, ou molécules, sont plus proches les unes des autres que la normale et ensuite plus éloignées les unes des autres. L'onde sonore se déplace vers l'extérieur du corps vibrant, mais les molécules n'avancent pas avec l'onde : elles vibrent autour d'un point de repos moyen. Les vibrations se produisent le long d'un axe qui est aligné avec la direction dans laquelle le son se propage. Cette forme d'onde est connue sous le nom d'"*onde longitudinale*". L'onde sonore s'affaiblit généralement au fur et à mesure que l'on s'éloigne du corps vibrant, et peut également subir des réflexions et des réfractions causées par les murs ou les objets situés sur son chemin, mais les molécules n'avancent pas avec l'onde : elles vibrent autour d'un point de repos moyen. L'onde sonore s'affaiblit généralement en s'éloignant de la source. Ainsi, l'"image" sonore atteignant l'oreille diffère quelque peu de celle initialement générée [24].

L'un des types de son les plus simples est l'onde sinusoïdale, également connue sous le nom de sinusoïde, dont la forme d'onde (variation de pression en fonction du temps) est illustrée à la figure 2.2. Cette onde est simple tant du point de vue physique et mathématique que du point de vue du système auditif. Il se trouve que les sinusoïdes produisent des réponses particulièrement simples dans le système auditif et qu'elles ont un son très propre ou "*pur*", comme celui d'un diapason. C'est pourquoi on les appelle aussi sons simples ou sons purs [24].

### 2.2.1.2 Caractéristiques du son

Pour décrire une sinusoïde, trois éléments doivent être précisés : *la fréquence*, ou le nombre de fois par seconde que la forme d'onde se répète (spécifié en Hertz, où 1 Hertz (Hz) = 1 cycle/s) ; *l'amplitude*, ou la quantité ou la variation de pression par rapport à la moyenne, et *la phase*, qui correspond à la partie du cycle par laquelle l'onde a avancé par rapport à un point fixe dans le temps. Lorsque l'onde parcourt un cycle complet, la phase change de  $360^\circ$ , ce qui équivaut à  $2\pi$  radians. Pour les sinusoïdes continues, la phase est seulement importante que lorsque l'on considère la relation entre deux ou plusieurs ondes différentes. Le temps nécessaire à un cycle complet de la forme d'onde s'appelle la période, qui est l'inverse de la fréquence. Par exemple, une sinusoïde de 500 Hz a une période de 2 ms [24].

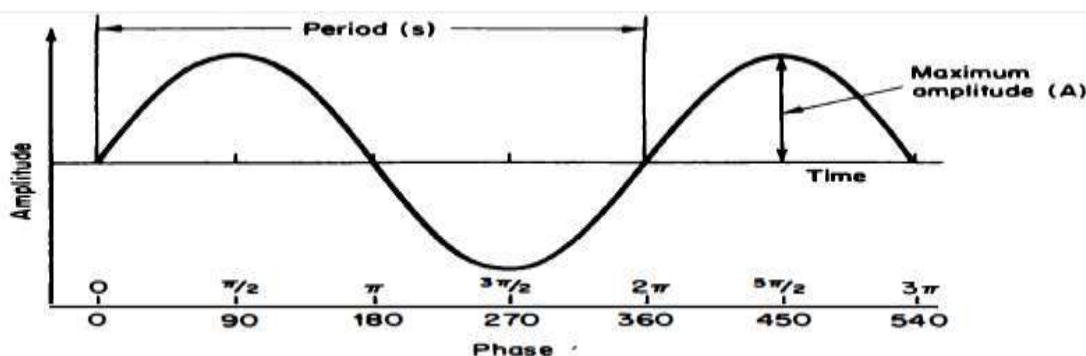


Fig. 2.2 - La forme d'une onde sinusoïdale ou vibration sinusoïdale.

Dans la figure 2.2, Seuls 1,5 cycles sont représentés. L'amplitude instantanée est donnée par l'expression  $A\sin(2\pi ft)$ , où  $t$  = temps,  $f$  = fréquence et  $A$  = amplitude maximale. La phase est indiquée le long du fond, en utilisant comme point de référence le premier passage par zéro de l'onde. La phase peut être mesurée en degrés ou en radians. Un cycle complet correspond à  $360^\circ$  ou  $2\pi$  radians.

Une onde sinusoïdale n'est pas le seul type de son qui se répète régulièrement. De nombreux sons de la vie quotidienne, comme ceux produits par les instruments de musique et certains sons de la parole, présentent également une telle régularité et, par conséquent, sont appelés sons périodiques. Bien que ces sons soient généralement plus complexes que les sinusoïdes, ils partagent une caractéristique subjective commune avec les sinusoïdes, en ce qu'ils ont une hauteur de son. L'hauteur peut être définie comme "*l'attribut de la sensation auditive en termes desquels les sons peuvent être classés sur une échelle allant du grave à l'aigu*" [25]. L'hauteur d'un son est liée à son taux de répétition et donc, dans le cas d'une sinusoïde, à sa fréquence (Figure 2.3).

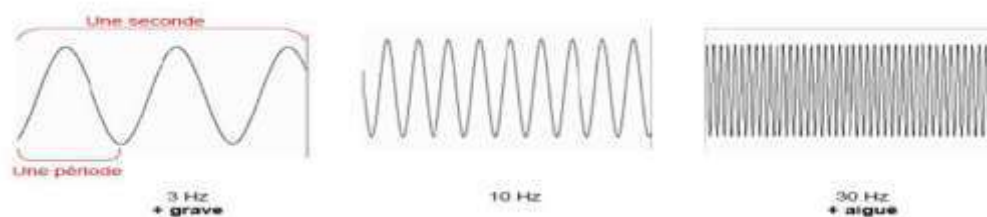


Fig. 2.3 - Les types de fréquence

On distingue les types de sons suivants :

- Son simple ou pur qui correspond à une vibration sinusoïdale unique (une seule fréquence), le son d'un diapason est un exemple d'un son simple.
- Son complexe périodique est un signal non sinusoïdal comme le son musical. Son spectre présente une fréquence fondamentale  $f_0$ , ainsi qu'un ensemble de fréquences  $f_n$ , dites harmoniques, multiples de la fréquence fondamentale :  $f_n = n \times f_0$ .
- Son complexe apériodique : à l'inverse, un son apériodique ne présente aucune périodicité comme son nom l'indique. Il possède une suite de plusieurs fréquences qui n'ont aucun lien entre elles, qui s'enchaînent sans aucune logique, de façon aléatoire et désorganisée. La parole ou le bruit sont des exemples de sons apériodiques. Le bruit blanc est un signal apériodique parfaitement aléatoire qui émet à intervalle irrégulière toutes les fréquences du spectre audible. Le bruit est un son jugé indésirable (Figure 2.4).

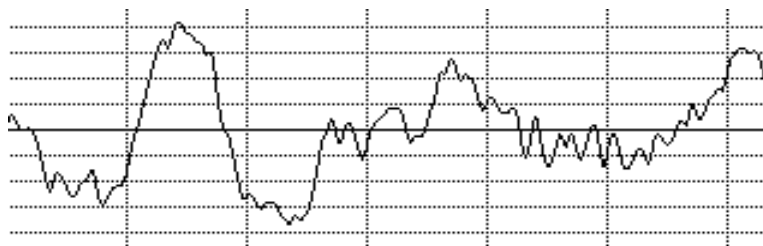


Fig. 2.4 - Amplitude en fonction du temps d'un signal apériodique

### 2.2.1.3 Analyse de Fourier et représentations spectrales

Bien que tous les sons puissent être spécifiés par leur variation de pression en fonction du temps il est souvent plus pratique, et plus significatif, de les spécifier d'une manière différente lorsque les sons sont complexes. Cette méthode est basée sur un théorème de Fourier, qui a prouvé que presque toutes les formes d'ondes complexes peuvent être analysées, ou décomposées, en une série de sinusoïdes avec des fréquences, des amplitudes et des phases spécifiques. Cela se fait à l'aide d'une procédure mathématique appelée la transformée de Fourier. L'analyse s'appelle l'analyse de Fourier, et chaque sinusoïde est appelée une composante (de Fourier) du son complexe. La forme d'onde complexe peut être produite en additionnant toutes les sinusoïdes qui la composent. Un son complexe peut donc être défini comme un son composé d'un certain nombre de sons simples, ou sinusoïdes [24].

Le type le plus simple de son complexe auquel on peut appliquer l'analyse de Fourier est celui d'un son périodique. Un tel son est composé d'un certain nombre de sinusoïdes, chacune d'entre elles ayant une fréquence qui est un entier multiple de la

fréquence d'une composante fondamentale commune. La composante fondamentale a donc la fréquence la plus basse de toutes les composantes du son complexe, et on peut dire qu'elle constitue la "base" des autres composantes. La composante fondamentale a une fréquence égale à la fréquence de répétition de la forme d'onde complexe dans son ensemble. Les composantes de fréquence du son complexe sont appelées harmoniques et sont numérotées, la fondamentale étant l'harmonique numéro 1 [24] (Figure 2.5).

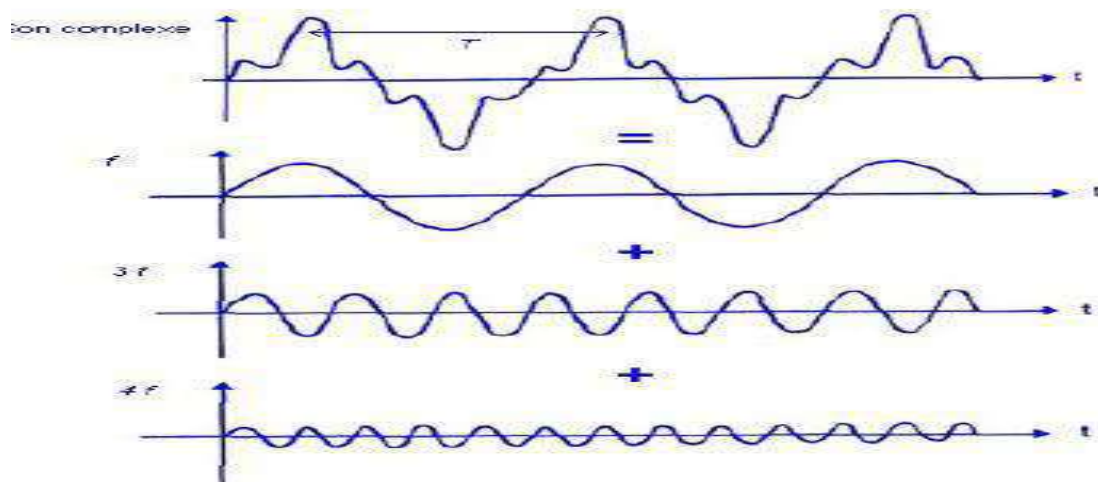


Fig. 2.5 - Analyse d'un son complexe

Ce sont les harmoniques qui donnent le timbre d'un son. Plus il y a d'harmoniques, plus le son sera agréable à l'écoute. Il est possible que deux signaux aient la même composante fondamentale, c'est alors la composition des harmoniques présentes au sein de chacun des deux sons qui vont permettre de les différencier (timbres différents). Ces spectres sont appelés spectres de raies (Figure 2.6).

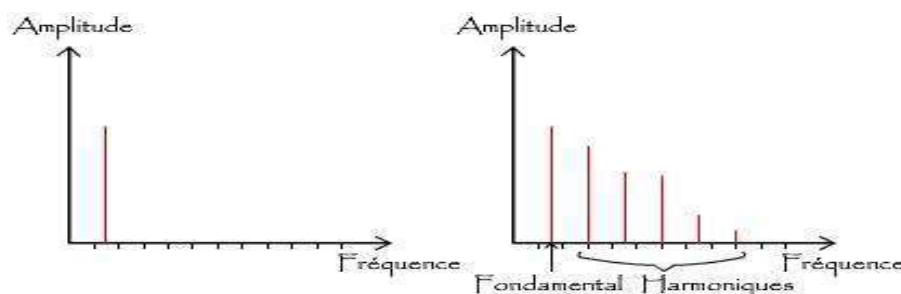


Fig. 2.6 - Spectre fréquentiel d'un son pur et d'un son complexe périodique

L'analyse spectrale d'un signal aperiodique (comme le bruit ou la parole par exemple) ne présente pas un spectre de raie (contrairement aux sons périodiques), mais un spectre continu. Cela s'explique par le fait que les sons aperiodiques possèdent une multitude de fréquences. Ces sons deviennent aléatoires lorsqu'ils présentent de variations de pression acoustique appelées "transitoires".



## 2.2.2 La mesure du niveau acoustique

Les instruments utilisés pour mesurer la magnitude des sons, tels que les microphones, réagissent normalement aux variations de la pression atmosphérique. Cependant, les magnitudes des sons sont souvent spécifiées en termes d'intensité, qui est l'énergie sonore transmise par seconde (c'est-à-dire la puissance). Pour un milieu tel que l'air, l'intensité est proportionnelle au carré de la variation de pression.

Il s'avère que le système auditif est capable de gérer une large gamme d'intensités sonores. Il est donc peu commode de traiter directement les intensités sonores. A la place, on utilise une échelle logarithmique exprimant le rapport entre deux intensités. Une intensité,  $I_0$ , est choisie comme référence et l'autre intensité,  $I_1$ , est exprimée par rapport à celle-ci. Un *Bel* correspond à un rapport d'intensités de 10 : 1. Ainsi, le nombre de *Bel* correspondant à un rapport d'intensités donné est obtenu en prenant le logarithme en base 10 de ce rapport d'intensité. Par exemple, un rapport d'intensité de 100 : 1 correspond à 2 Bels. Le Bel correspond à un rapport d'intensités assez important pour l'usage quotidien, et pour obtenir des unités de taille plus pratique, le Bel est divisé en 10 *décibels* (*dB*). Ainsi, le nombre de décibels correspondant à un rapport donné d'intensité acoustique est le suivant [24] :

$$\text{Nombre de décibels} = 10 \log_{10}(I_1 / I_0) \quad (1.1)$$

Lorsque la magnitude d'un son est spécifiée en décibels, il est habituel d'utiliser le mot "niveau" pour se référer à sa magnitude. Il faut noter qu'un nombre donné de décibels représente une intensité ou un rapport de puissance, et non une intensité absolue. Pour spécifier l'intensité absolue d'un son, il est nécessaire d'indiquer que l'intensité du son,  $I_1$ , est supérieure ou inférieure de  $X$  dB à une intensité de référence,  $I_0$ . L'intensité de référence la plus couramment utilisée, pour les sons dans l'air, est de  $10^{-12} \text{ W/m}^2$  (watts par mètre carré), ce qui équivaut à une pression de  $2 \times 10^{-5} \text{ N/m}^2$  ou  $20 \mu\text{Pa}$  (micropascal). Un niveau sonore spécifié à l'aide de ce niveau de référence est appelé niveau de pression acoustique (SPL-Sound Pressure Level) ou niveau d'intensité sonore. Ainsi, un son d'un SPL de 60 dB est 60 dB de plus par rapport au niveau de référence de 0 dB et a une intensité de  $10^{-6} \text{ W/m}^2$ . Notez qu'en multipliant (ou en divisant) le rapport des intensités par 10 augmente (ou diminue) le nombre de décibels de 10. Il est également utile de se rappeler qu'une variation d'intensité de deux fois plus élevée correspond à une variation de niveau de 3 dB (car le logarithme de 2 est approximativement égal à 0,3) [24].

Le niveau sonore de référence, de 0 dB, est un niveau sonore qui a été choisi pour être proche du seuil absolu moyen des humains pour une sinusoïde de 1000 Hz. Le seuil absolu est le niveau minimum détectable d'un son en l'absence de tout autre son externe. En fait, le seuil absolu moyen à 1000 Hz est d'environ 6.5 dB (en écoutant avec une seule oreille).

Enfin, il est utile d'adapter la notation en décibels pour qu'elle exprime des rapports de pression ainsi que des rapports d'intensité. Cela peut se faire en rappelant que l'intensité est proportionnelle au carré de la pression. Si un son a une intensité de  $I_1$  et une pression  $P_1$  et qu'un second son a une intensité de  $I_2$  et une pression  $P_2$ , alors la différence de niveau entre eux est [24]:

$$SPL (dB) = 10 \log_{10}(I_1 / I_2) = 10 \log_{10}\left(\frac{P_1}{P_2}\right)^2 = 20 \log_{10}(P_1 / P_2) \quad (1.2)$$

Ainsi, une augmentation de 10 fois la pression correspond à une augmentation de 100 fois de l'intensité et est représentée par +20dB. Le tableau 2.1 [24] donne quelques exemples de rapports d'intensité et de pression, et indique également les niveaux sonores, SPL en dB, correspondant à divers sons courants.

Tableau 2.1 - La relation entre les décibels, les rapports d'intensité et les rapports de pression.

Niveau du son (SPL en dB)	rapport d'intensité ( $I_1/I_2$ )	rapport de pression ( $P_1/P_2$ )	Exemple typique
140	$10^{14}$	$10^7$	Coup de feu à bout portant
120	$10^{12}$	$10^6$	Groupe de rock bruyant
100	$10^{10}$	$10^5$	Crier à bout portant
80	$10^8$	$10^4$	Rue animée
70	$10^7$	3160	Conversation normale
50	$10^5$	316	Conversation discrète
30	$10^3$	31.6	Un doux murmure
20	$10^2$	10	Paysage de nuit
6.5	4.5	2.1	Seuil absolu moyen à 1 kHz
3	2	1.4	
0	1	1	Niveau de référence
-10	0.1	0.316	

Les niveaux sonores SPL en dB sont exprimés par rapport à une intensité de référence  $I_0$  de  $10^{-12}$  W/m<sup>2</sup>. Ceci est équivalent à une pression de 20 µPa.

### 2.2.3 Numérisation d'un signal audio analogique

Un signal audio analogique est une forme d'onde électrique continue dans le temps. Un convertisseur Analogique/ Numérique traduit ce signal en séquences de nombres binaires. Le processus de conversion A/N a une incidence majeure sur la qualité finale du signal audio numérique. En effet la qualité d'un signal audio une fois converti ne peut jamais s'améliorer.

Les principaux paramètres influant sur la qualité audio sont le taux d'échantillonnage et le nombre de bits par échantillon. Le convertisseur de base intégré à l'ordinateur multimédia est un convertisseur 16 bits/44,1 kHz. Le convertisseur très haut de gamme est plutôt un convertisseur 24 bits/96 kHz.

#### 2.2.3.1 L'échantillonnage du son

Pour pouvoir représenter un son sur un ordinateur, il faut le convertir en valeurs numériques, car celui-ci ne sait travailler que sur ce type de valeurs. Il s'agit donc de relever des petits échantillons de son (ce qui revient à relever des différences de

pression) à des intervalles de temps précis (Figure 2.7). On appelle cette action l'échantillonnage ou la numérisation du son. L'intervalle de temps entre deux échantillons est appelé taux d'échantillonnage.

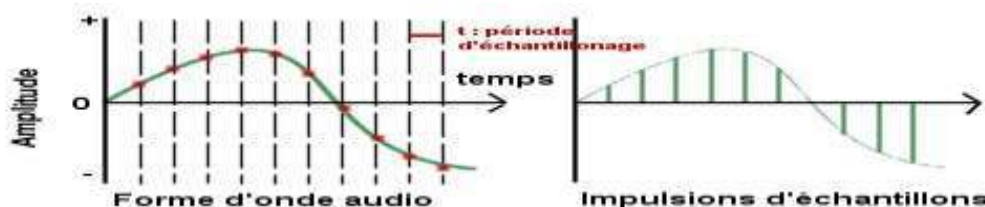


Fig. 2.7 - Echantillonnage du son

Etant donné que pour arriver à restituer un son qui semble continu à l'oreille il faut des échantillons tous les quelques 100000ème de seconde, il est plus pratique de raisonner sur le nombre d'échantillons par seconde, exprimés en Hertz(Hz). Voici quelques exemples de taux d'échantillonnage et de qualités de son associées (Tableau 2.2).

Tableau 2.2 - Taux d'échantillonnage de certaines applications audio

Taux d'échantillonnage	Qualité du son
44100 Hz	qualité CD
22000 Hz	qualité radio
8000 Hz	qualité téléphone

La valeur du taux d'échantillonnage, pour un CD audio par exemple, n'est pas arbitraire, elle découle en réalité du théorème de Shannon. La fréquence d'échantillonnage doit être suffisamment grande, afin de préserver la forme du signal.

Le Théorème de Nyquist – Shannon stipule que la fréquence d'échantillonnage doit être égale ou supérieure au double de la fréquence sonore pour numériser correctement un son. Notre oreille perçoit les sons environ jusqu'à 22000 Hz, il faut donc une fréquence d'échantillonnage au moins de l'ordre de 44000 Hz pour obtenir une qualité satisfaisante. Il existe un certain nombre de fréquences d'échantillonnage normalisées :

- 32 kHz : pour la radio FM en numérique
- 44.1 kHz : pour l'audio professionnelle et les compact-disques
- 48 kHz : pour les enregistreuses numériques multipistes professionnelles

### 2.2.3.2 La quantification

Quantifier un signal consiste à placer les amplitudes des échantillons sur une échelle de valeurs à intervalles fixes. Le quantificateur détermine dans quel intervalle de quantification (Q) l'échantillon se situe et à pour but de lui affecter une valeur représentant le point central de cet intervalle. Chaque amplitude possède donc un nombre binaire unique (Figure 2.8).

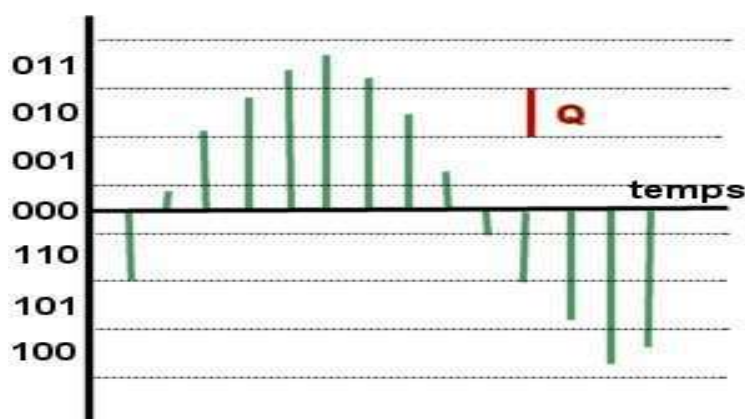


Fig. 2.8 - Valeurs binaires des intervalles de quantification (quantificateur 3 bits)

Dans un système binaire, le nombre de pas de quantification est égal à  $2^n$  avec  $n$  le nombre de bits utilisés par mot binaire. Donc un mot binaire de 4 bits offre 16 niveaux de quantification alors qu'un quantificateur 16 bits en offre 65536 niveaux. Avec la seconde représentation, on aura bien évidemment une qualité de son bien meilleure, mais aussi un besoin en mémoire beaucoup plus important.

Enfin, la stéréophonie nécessite deux canaux sur lesquels on enregistre individuellement un son qui sera fourni au haut-parleur de gauche, ainsi qu'un son qui sera diffusé sur celui de droite. Si les signaux dépassent la limite fixée par le nombre de bits utilisés, alors cela induit une distorsion importante car ces signaux sont tout simplement écrêtés car il n'existe pas de valeurs possibles pour les représenter. L'erreur de quantification est considérée comme un signal indésirable ajouté au signal utile. Ces signaux sont qualifiés de distorsion ou de bruit.

### 2.2.3.3 Modulation par impulsion codée PCM (Pulse Code Modulation)

La modulation par impulsion codée est un format standard pour le stockage ou la transmission de données numériques non compressées. Les CDs, les unités DATs sont quelques exemples de supports qui adaptent le format PCM. Il existe deux paramètres pour le PCM : la fréquence d'échantillonnage [Hz] et le débit binaire ou bit-rate [Bit]. La fréquence d'échantillonnage décrit le nombre d'échantillons par seconde de l'enregistrement. Un taux d'échantillonnage élevé implique que des fréquences plus élevées seront incluses. Le bit-rate décrit la taille du mot binaire qui contient la valeur de l'échantillon. Un bit-rate plus élevé donne une meilleure résolution audio et un bruit plus faible car l'échantillon peut être déterminé plus précisément en utilisant plus de bits. Les CD audio ont une fréquence de 44 100 Hz et un débit de 16 bits [26]. Une façon rudimentaire de compresser l'audio serait de l'enregistrer à une fréquence d'échantillonnage ou un débit binaire inférieur. L'utilisation d'un débit binaire de 8 bits au lieu de 16 bits réduira la quantité de données de 50 %, mais la perte de qualité qui en résulte est inacceptable.

## 2.2.4 La compression du signal audio numérisé

Un signal audio numérisé est stocké sur des disques durs, des disques compacts, des DVD... La nature de l'information qu'ils contiennent rend ces fichiers relativement volumineux.

Le son numérique s'est généralisé au travers du CD audio (PCM linéaire). Un CD audio peut contenir en moyenne 74 minutes de musique s'il utilise 44100 échantillons de 16 bits par seconde. La capacité totale du CD est de 650 Mo environ. Ce résultat est tout à fait correct, et a permis un grand développement au CD audio. En revanche, il est très difficile de transmettre ce son par un canal : il faut un débit de 705.6 KBits/s pour jouer un CD mono et 1.411 MBits/s pour la stéréo. Difficile même avec l'arrivée d'internet à haut débit.

Ces débits élevés induisent des coûts élevés que ce soit en termes de stockage ou de transfert de données. L'objectif de la compression du son est donc de réduire considérablement la taille des données tout en essayant de garder une qualité similaire à l'original. Comme la compression repose sur la suppression de données, l'enjeu est de déterminer avec précision quelles données à supprimer sans détériorer le signal.

Les techniques de réduction de débit sont déjà très largement employées dans les domaines du cinéma et de la radio, via le câble, le satellite.

### 2.2.4.1 Les algorithmes de compression

Un algorithme est l'énoncé d'une suite d'opérations permettant de donner la réponse à un problème. Dans le cas de la compression, l'algorithme a pour fonction de réduire la taille d'un fichier selon un certain nombre de contraintes que le programmeur spécifie. Par exemple, une des contraintes peut être de conserver toutes les fréquences inférieures à 20kHz afin de limiter les pertes de qualité sonore dans la zone audible du spectre. Lors de l'étape de compression et de décompression d'un flux audio ou vidéo, on utilisera des algorithmes spécifiques rassemblés sous le terme commun de "*CoDec*" (COde-DECode en anglais). Un codec est constitué de deux éléments :

- le *COdeur* contient un algorithme destiné à coder l'information. Dans le cas de la compression il permet d'effectuer une réduction du poids des données ;
- le *DECodeur* contient un algorithme destiné à décoder l'information. Dans le cas de la compression il permet de reconstruire un signal audio numérique.

### 2.2.4.2 Le taux de compression

Compresser revient à réduire le débit du flux audio et/ou vidéo. Les algorithmes sont adaptés en fonction des applications (diffusion internet, télévision, cinéma) pour répondre aux besoins de chacun des médias. La réduction de débit (ou compression) s'exprime généralement sous la forme d'un taux dit "*taux de compression*". Le taux de compression peut s'énoncer comme suit :

- Soit comme le rapport entre le volume initial des données et le volume après réduction. Si le volume de données est deux fois plus faible après réduction (passant de 10Mo à 5Mo par exemple), on écrira qu'il s'agit d'un taux de 2:1 ;

- Soit en pourcentage du volume après réduction par rapport au volume initial. Si le volume de données est deux fois plus faible après réduction, on écrira qu'il s'agit d'un taux de 50%. Il existe par ailleurs deux types de compressions : la compression " destructive ou avec perte" et la compression "non destructive ou sans perte".

## 2.3 FORMAT DE FICHER AUDIO

Un format de fichier audio est un format de données utilisé en informatique pour stocker des sons, (de la musique, des voix, etc.) sous forme numérique. Plusieurs standards existent; certains s'appliquent à la production, au stockage et à la diffusion, d'autres (ceux qui utilisent des algorithmes de compression de données ou de réduction de débit), sont destinés, en principe, uniquement à la diffusion [27]. Le format audio non compressé le plus utilisé en production ou stockage est le PCM qui est stocké sous forme WAF sur Windows ou sous AIFF sur MAC OS. Le format compressé le plus utilisé en diffusion est de loin le *MP3*, suivi du *WMA*, et de l'*AAC*.

### 2.3.1 Caractéristique des formats audio

Les formats audio varient selon :

- Le nombre de canaux sonores encodés.
- Le nombre d'échantillons par seconde avec lequel on découpera numériquement, pour chaque canal, une onde sonore ou un signal électrique.
- La résolution donnée à chaque échantillon et la grandeur physique qu'on lui donne.
- L'utilisation d'une compression ou non.

Chaque format audio présente aussi des caractéristiques découlant de l'algorithme de compression/décompression, ou *CODEC*, qu'il utilise ou non. Après la numérisation du son, le format utilisé est inscrit dans l'extension du fichier de données qui en stocke la transcription. Chaque format se caractérise aussi par sa disposition à inclure et gérer des Métadonnées.

Dans un format donné, les fichiers peuvent être déclinés en plusieurs échelles de quantification (8, 16, 24 ou 32 bits) avec différentes fréquences d'échantillonnage (par exemple 22.05, 44.1, 48, 88.2 , 96, 176.4, 192, kilohertz) appliqués à un certain nombre de voies (monophonique, stéréophonique).

Le nombre de canaux sonores peuvent être réels et séparés, ou mélangés discrètement aux signaux principaux, ils seront décodés et restitués par la suite à l'aide d'algorithmes spécifiques.

### 2.3.2 Les formats audio sans compression

Le format audio non compressé est le PCM, qui est généralement stocké sous forme de WAV sur Windows ou sous AIFF sur Mac OS [27]. WAV et AIFF sont des formats de fichiers flexibles conçus pour stocker plus ou moins n'importe quelle combinaison de taux d'échantillonnage ou de débit binaire (bit rate). Ce sont les formats de fichier appropriés pour le stockage et la réalisation d'enregistrements originaux. Le tableau 2.3 décrit les principaux formats non compressés de fichiers audio.

Tableau 2.3 - Formats audio non compressés

<i>Format</i>	<i>Description</i>
<b>RIFF</b> (Resource Interchange File Format)	Plusieurs formats audio étant basés sur la norme RIFF, format d'échange de fichiers ressources, il s'agit d'une norme d'usage général
<b>WAV</b> ou WAVE, (WAVEform audio format)	C'est un conteneur capable de recevoir des formats variés. Il est basé sur le format de fichier RIFF lequel est semblable au format IFF. Mono ou stéréo, il a été mis au point par Microsoft et IBM [28]. Le suffixe des fichiers créés est .wav. La taille de ce fichier dépend de la fréquence d'échantillonnage, du type de son (mono ou stéréo) et du nombre de bits utilisés pour l'échantillonnage (8 ou 16 bits).
<b>AIFF</b> ( <i>Audio Interchange File Format</i> )	C'est un format de stockage de sons sur les ordinateurs de Apple Mac OS. C'est l'équivalent du format WAV dans le monde Windows. Les résolutions 8, 16, 20, 24 et 32 bits (à virgule flottante) sont acceptées. Le suffixe des fichiers créés est .aif ou .aiff. Les données sont codées en PCM big-endian sans compression. Ainsi, une piste CD Audio, codée en 16 bits, stéréo 44,1 kHz aura un bitrate de 1.411Mbit/s. Il est devenu le 1er format d'échange multiplateformes. Il existe néanmoins un format compressé (AIFF-C ou AIFC) qui supporte une compression pouvant aller jusqu'à un rapport 1/6 [29].

Autres formats non compressés existent comme **RAW** (Real Audio Wrapper), **BWF** (*Broadcast Wave Format*) [30], **CAF** (*Core audio format*) [31], et **CDA** (*Compact Disc Audio*) [32].

### 2.3.3 Les formats audio compressés sans perte

L'objectif principal de la compression du son est de conserver une qualité sonore maximum tout en réduisant le plus possible la taille des données. La compression de données fonctionne en supprimant l'information inutile ou redondante du flux de données, tout en préservant la qualité sonore.

La compression sans perte (*lossless*) ou non destructive signifie qu'on utilise un algorithme tel qu'on peut toujours retrouver les données d'origine. Des algorithmes spécifiques comme WavArc, RKAU, Perfect Clarity Audio (\*.wa, \*.rka, \*.pca) sont spécifiquement dédiés au son et non destructifs, et permettent d'atteindre une réduction de volume de 50%.

Typiquement, la compression sans perte permet de diviser la taille des fichiers par deux ou trois. Elle est relativement peu utilisée, car ce gain est très faible en comparaison de ceux permis par la compression avec perte [27]. Ces algorithmes nécessitent plus de temps de calcul et sont moins appropriés pour les échanges de fichiers. Aucun standard n'a donc suffisamment convaincu pour devenir universellement utilisable. Le tableau 2.4 décrit les deux principaux formats audio compressés sans perte.

Tableau 2.4 – Formats audio compressés sans perte

<i>Format</i>	<i>Description</i>
<b>ATRAC</b> ( <i>Adaptive Transform Acoustic Coding</i> )	Une technique de compression audio avec et sans pertes développée par Sony en 1992. Ce format a subi plusieurs évolutions : ATRAC3, ATRAC3 plus (familièrement écrit ATRAC3+) et ATRAC Advanced Lossless se sont succédé respectivement en 1999, 2002 et 2006 [33]. Seul l'ATRAC Advanced Lossless propose une compression sans perte. Les extensions de ce fichier sont .aa3, .oma, et .at3 [34].
<b>FLAC</b> ( <i>Free Lossless Audio Codec</i> )	Un format audio similaire au MP3, mais sans perte, ce qui signifie que l'audio est compressé en FLAC sans aucune perte de qualité. Ce format est maintenu par la fondation Xiph.org. Le fonctionnement est similaire à celui de Zip, sauf qu'avec le FLAC, on obtiendra une bien meilleure compression car il est conçu spécifiquement pour l'audio. Nous pouvons lire les fichiers FLAC compressés dans tout lecteur préféré comme nous pouvions le faire avec un fichier MP3. Le FLAC se distingue comme étant le codec audio sans perte le plus rapide et le plus largement supporté, et le seul qui à la fois n'est pas propriétaire, n'est pas encombré de brevets, a une implémentation de référence open-source, a un format et une API bien documentés, et a plusieurs autres implémentations indépendantes [35].

### 2.3.4 Les formats audio compressés avec perte

La compression audio avec perte (*lossy*) se base sur des algorithmes spécialisés pour déterminer quelles transformations simplifient la représentation du son tout en étant perçue en entier de la même manière par l'oreille humaine. Elle diminue la taille du fichier en éliminant les nuances perçues comme les moins utiles. L'élimination est définitive, créer un fichier dans un format de haute qualité à partir d'un fichier compressé avec perte ne sert strictement à rien.

Le format le plus connu est le *MPEG-1/2 Audio Layer 3*, dont le suffixe est .mp3. Ce format propose une qualité sonore très correcte pour un débit de 128 kbit/s. C'est ce format qui a été massivement utilisé pour transférer les musiques via internet dès la fin des années 1990. Rapidement, des *baladeurs* avec une mémoire réenregistrable et capables de lire directement ce format sont apparus.

Dans la décennie 2000, de nouveaux formats ont été proposés. Vu les progrès des algorithmes, ils surpassent largement le MP3 en termes de qualité à débit égal, et peuvent atteindre des qualités supérieures. De plus, certains sont moins contraignants que le MP3 quant aux droits d'utilisation (l'Ogg est un format libre). Mais le MP3 reste le plus utilisé, car l'arrivée en continu de nouveaux formats, apportant un avantage assez faible par rapport aux précédents, ne permet pas de mettre en place un standard meilleur que le MP3 et lisible par tous les baladeurs. En particulier, le fait que les iPod d'Apple ne lisent que le MP3 et l'AAC (*Advanced Audio Coding*) freine assez fortement les initiatives dans ce domaine.



Pour un même format de compression, il n'y a pas de manière unique de coder, car chaque algorithme cherche la meilleure manière de représenter le son d'origine suivant le langage de compression. En particulier, les codecs de MP3 ont réalisé des progrès très importants depuis le début de l'utilisation de ce format. Ils permettent typiquement un gain d'un facteur de 1:10 [27]. Cela a rendu possible non seulement le stockage d'un temps d'écoute formidable sur les supports informatiques, mais aussi leur échange par Internet, souvent illégalement.

La compression MP3 utilise le codage perceptuel, une approche basée sur des modèles psycho-acoustiques qui permettent au codec d'écarter ou de réduire la précision des composants audio qui sont moins audibles pour l'oreille humaine. Les trois classes de compression audio associées aux spécifications MPEG-1 et -2 sont connues sous le nom de couches I, II et III, MP3 est l'abréviation de la couche III. Un fichier MP3 créé avec un débit binaire de 128 kbit/s est environ 1/11 de la taille d'un fichier LPCM non compressé aux niveaux de qualité d'un disque compact (44,1 kHz, 16 bits de profondeur). Un fichier MP3 peut également être enregistré à des débits binaires supérieurs ou inférieurs, avec une qualité résultante supérieure ou inférieure [36]. Les formats audio compressés avec perte sont décrits dans le tableau 2.5.

Tableau 2.5 – Formats audio compressés avec perte

<b>Format</b>	<b>Description</b>
<b>AC3</b> ( <i>Audio Coding 3</i> )	Officialisée en 1992, Il s'agit d'une extension pour les fichiers audio de son surround. Elle a été inventée en 1987 par les Laboratoires Dolby afin d'être utilisée sur les DVD, les lecteurs Blu-ray, et les systèmes domestiques de programmation et de divertissement en haute définition (HDTV). La compression AC3 permet d'utiliser jusqu'à 6 canaux sonores indépendants avec un taux d'échantillonnage de 32, 44,1 ou 48 kHz et avec un taux de transfert allant de 32 à 640 kbit/s. Le Dolby Digital utilise ce principe de codage, c'est pourquoi on le désigne souvent sous ce nom. Format très courant dans les DVD [37].
<b>MP3</b> (MPEG-1/2 Audio Layer 3)	Cet algorithme de compression prend naissance en 1987. Conçu conjointement par l'institut Fraunhofer et l'entreprise Thomson, et standardisé par l'ISO dès 1992, le MP3 est un dérivé de la compression sonore utilisée pour les vidéos au format MPEG-1 (MP3 signifie d'ailleurs MPEG-1 Audio Layer 3). La couche (Layer) 3 est la couche la plus complexe. Elle est dédiée à des applications nécessitant des débits faibles (128 kbit/s) d'où une adhésion très rapide du monde Internet à ce format de compression. Les taux de compression (ratio) sont d'ordinaire de 1 pour 10 (1:10) (1:4 à 1:12). Très rapide à l'encodage. Des royalties importantes sont à payer pour exploiter la licence MP3 [38].
<b>mp3PRO</b>	Le format mp3PRO, fruit de la collaboration entre Thomson Multimédia et l'Institut Fraunhofer, combine l'algorithme MP3 et un système améliorant la qualité des fichiers comprimés appelé SBR pour Spectral Bandwidth Replication. Ce format a été publié à la fin de 2001, un fichier MP3pro 64 kbit/s a une qualité équivalente à celle d'un MP3 à 128 kbit/s. un fichier mp3Pro permet un gain de place substantiel par rapport à un fichier mp3 standard (entre 25% et 50% de gain) [39]. . Le suffixe des fichiers créés est .mp3.

<p><b>WMA</b> (<i>Windows Media Audio</i>)</p>	<p>Lancé en 1999 par Microsoft, le WMA n'a eu pour but que de contrer l'essor du MP3, avec des arguments convaincants. Exploitant plus efficacement les caractéristiques de l'oreille humaine au niveau du spectre audible, le WMA s'attache à supprimer les fréquences réellement inutiles tout en conservant certaines hautes fréquences absentes avec un MP3. Il existe des variantes "lossless" de ce format, notamment le WMA 10 Pro capable de gérer des flux audio Hi-Res 24 bits 96 kHz avec un débit pouvant atteindre 768 kbps. Autre argument à destination des professionnels, le WMA offre lui aussi une gestion pointue des droits d'auteur (DRM ou <i>Digital Right Management</i>). Il est possible de définir par exemple une durée de vie limitée pour les fichiers ou interdire leur copie.</p>
<p><b>Ogg Vorbis</b></p>	<p>C'est un format libre, fruit de la fondation Xiph.org. Vorbis est un algorithme de compression et de décompression (codec) audio numérique, sans brevet, ouvert et libre, plus performant sur le plan de la qualité et du taux de compression que le format MP3, mais moins populaire que ce dernier. Il est remplacé dans les nouvelles applications par le codec Opus. Le format sonore Vorbis doit obligatoirement être encapsulé. Il l'est dans la majorité des cas dans le format Ogg, d'où l'appellation de Ogg Vorbis. Néanmoins, Vorbis et Ogg sont différents (l'un étant le contenu et l'autre le contenant). Ogg Vorbis est donc le format audio Vorbis encapsulé dans le contenant Ogg [40].</p>
<p><b>AU</b></p>	<p>C'est un format assez bien répandu grâce à Unix et Linux. Le fichier AU fut introduit par Sun Microsystems. Ce format facilite le stockage de fichiers audio, généralement en trois parties : un en-tête composé d'un maximum de 24 octets, un bloc d'annotation de longueur variable et les données audio elles-mêmes. Ce format de fichier est essentiellement utilisé par Sun et des machines basées sur Unix, mais il peut être ouvert en utilisant différents programmes audio [41]. Le suffixe des fichiers créés est .au.</p>

## 2.4 CODAGE AUDIO PERCEPTIF UTILISANT DES MODELES PSYCHO-ACOUSTIQUES

Au cours des deux dizaines dernières années, nous avons assisté à une croissance phénoménale de la technologie audio numérique, qui constitue une partie essentielle des normes et de la technologie multimédia. L'audio numérique trouve de nombreuses applications dans des domaines tels que le stockage de CD/DVD, la téléphonie numérique, la diffusion par satellite, l'électronique grand public, etc. Dans cette section, nous présentons les concepts de l'encodage audio basés sur les modèles psycho-acoustiques car c'est ce type qui s'adapte mieux au type de fichier audio et s'appuie sur les caractéristiques du système auditif humain (HAS- Human Auditory System).

### 2.4.1 Les classes du signal audio

En fonction des applications, les signaux audio peuvent être regroupés en trois grandes classes [42], [43]:

**La voix téléphonique** : Il s'agit d'une application à faible bande passante, couvrant la gamme de fréquences de 300 à 3400 Hz. L'intelligibilité et le naturel de la voix, bien

que médiocres, sont tout juste acceptables pour la téléphonie et certains services de vidéo téléphonie.

**Voix à large bande** : Elle couvre une bande passante de 50Hz à 7000Hz pour une meilleure qualité de la voix.

**Audio à large bande** : Il s'agit d'applications audio haute fidélité (voix, ainsi que musique) nécessitant une bande passante d'au moins 20 KHz pour le stockage audio numérique et les applications de diffusion.

Le format numérique conventionnel de ces signaux est le PCM, avec des taux d'échantillonnage et des résolutions d'amplitude (bits PCM par échantillon) comme indiqué dans le tableau 2.6.

Tableau 2.6 - Paramètres de base des classes du signal audio acoustique

Classe du signal audio	Intervalle de fréquence en Hz	Taux d'échantillonnage en kHz	Bits PCM par échantillon	Débit binaire PCM en kb/s
Voix téléphonique	300 - 3400 <sup>1</sup>	8	8	64
Voix à large bande	50 - 7000	16	8	128
Audio à large bande (stéréo)	10 - 20000	48 <sup>2</sup>	2 x 16	2 x 768

Pour l'archivage et le traitement des signaux audio, des taux d'échantillonnage d'au moins 2 x 44.1 kHz et des résolutions d'amplitude jusqu'à 24 bits par échantillon sont utilisés. Le codage sans perte est très important pour ne pas compromettre la qualité audio. Le disque numérique polyvalent (DVD), avec sa capacité de 4,7 Go est le support de stockage approprié pour de telles applications.

## 2.4.2 Débit binaire requis pour l'audio de qualité stéréo

Dans les premières années de la technologie audio numérique, la qualité audio stéréo des disques compacts (CD) était utilisée comme norme avec les spécifications suivantes :

- Fréquence d'échantillonnage : 44.1 KHz
- 16 bits/échantillon pour chacun des deux canaux stéréo.

Par conséquent, le débit binaire net requis est de  $2 \times 16 \times 44,1 \times 1000 = 1.41$  Mbits/s (Tableau 2.6). Cependant, des bits supplémentaires sont nécessaires pour la synchronisation et la correction des erreurs, ce qui donne 49 bits pour chaque échantillon audio de 16 bits. Ainsi, le débit binaire stéréo total requis est de  $1.41 \times 49/16$  Mbit/sec = 4.32 Mbit/sec. Le tableau 2.7 compare les débits binaires du disque compact et de la bande audio numérique (DAT) [42], [43].

Bien que des canaux à large bande passante soient disponibles, il est nécessaire de réaliser une compression pour les applications à faible débit binaire dans le cadre d'un stockage et d'une transmission rentables. Dans des applications telles que la radio

<sup>1</sup> Large bande en Europe, 200 à 3200 Hz en USA

<sup>2</sup> Autres taux d'échantillonnage 44.1 kHz, 32 kHz

mobile, les canaux ont une capacité limitée et une compression efficace de la bande passante doit être utilisée.

Tableau 2.7 - Débits binaires CD et DAT

<i>Dispositif du stockage</i>	<i>Débit audio (Mb/s)</i>	<i>Overhead (Mb/s)</i>	<i>Total débit binaire (Mb/s)</i>
Compact Disc (CD)	1.41	2.91	4.32
Digital Audio Tape (DAT)	1.41	1.05	2.46

*Note : Signaux stéréophoniques, échantillonnés à 44.1 KHz, DAT supporte aussi des taux d'échantillonnage de 32 kHz et 48 kHz*

### 2.4.3 Portée des normes audio MPEG

Le MPEG (*Moving Pictures Experts Group*) a fourni les normes pour le codage audio numérique, dans le cadre des normes multimédia. Trois normes, à savoir MPEG-1, MPEG-2 et MPEG-4, répondent aux besoins suivants [42], [43] :

**a) MPEG-1 audio** : Dans cette norme, sur un débit total de 1.5 Mbits/s pour un stockage multimédia de qualité CD, 1.2 Mbits/s est alloué à la vidéo et 256 Kbits/s à l'audio. Jusqu'à deux canaux audio sont pris en charge.

**b) MPEG-2 audio** : Cette norme répond aux exigences des applications HDTV (*High Definition TV*). Dans sa partie audio, deux à cinq canaux audio à bande passante complète sont pris en charge. La norme offre également des outils connus sous le nom de codage audio avancé (*MPEG-2 AAC*).

**c) MPEG-4 audio** : Les normes MPEG-4 pour le codage audiovisuel s'adressent à des applications allant de l'accès mobile, des terminaux multimédia peu complexes aux systèmes sonores multicanaux très complexes. La principale caractéristique du MPEG audio est qu'au lieu d'utiliser un modèle pour la source audio (comme le modèle du tractus vocal utilisé pour les signaux vocaux), les codeurs exploitent les limites perceptives du système auditif humain. La compression est réalisée en éliminant les parties de l'audio non pertinentes sur le plan perceptif, qui ne peuvent produire aucune distorsion audio.

### 2.4.4 Les technologies clés du codage audio

Les premières propositions visant à réduire les taux de codage audio à large bande ont suivi celles du codage de la voix. Les différences entre les signaux audio et les signaux vocaux sont multiples. Le signal vocal est un cas particulier de signaux audio, et il est différent des signaux audio typiques en termes de largeur de bande spectrale, de distribution de l'intensité et de continuité du signal [44], les techniques développées pour le tatouage des fichiers audio typiques ne sont pas toujours applicables à la parole. Cependant, le codage audio implique des taux d'échantillonnage plus élevés, une meilleure résolution d'amplitude, une gamme dynamique plus étendue, des variations plus importantes des spectres de densité et de puissance, des présentations de signaux audio stéréophoniques et multicanaux et, enfin, des attentes plus élevées de l'auditeur en matière de qualité. En effet, la haute qualité du CD avec son format PCM à 16 bits par échantillon a rendu l'audio numérique populaire [42].

Le codage de la voix et le codage audio sont similaires dans la mesure où, dans les deux cas, la qualité est basée sur les propriétés de la perception auditive humaine. D'autre part, la voix peut être codée très efficacement car un modèle de production de

la parole est disponible, alors que rien de semblable n'existe pour les signaux audio [42].

Des réductions modestes des débits binaires audio ont été obtenues par le compactage instantané (par exemple, une conversion d'un PCM 14 bits uniforme en une présentation PCM non uniforme de 11 bits) ou par un PCM adaptatif à l'avance (compactage par blocs) tel qu'il est employé dans diverses formes de codage multiplex audio à compactage quasi instantané (NICAM) [UIT-R, Rec. 660] [42]

Des réductions de débit binaire par des moyens assez simples sont réalisées dans le CD interactif (CD-i) qui prend en charge les éléments suivants : PCM 16 bits à une fréquence d'échantillonnage de 44.1 kHz et permet trois niveaux de PCM différentiel adaptatif (ADPCM) avec prédiction commutée et mise en forme du bruit. Les bandes passantes et les résolutions d'échantillonnage prises en charge sont les suivantes 37.8 kHz/8 bits, 37.8 kHz/4 bits et 18.9 kHz/4 bits [42].

Dans les algorithmes de codage audio récents, quatre technologies clés jouent un rôle important : le codage perceptuel, le codage dans le domaine fréquentiel, la commutation de fenêtre et l'allocation dynamique de bits [42].

#### **2.4.4.1 Un aperçu sur les techniques du codage perceptuel**

Les modèles psycho acoustiques de la perception auditive humaine ont trouvé une application importante dans le domaine du codage audio perceptif, où l'exploitation des limites de la perception et la réduction de la non-pertinence sont essentielles pour obtenir une réduction significative du débit binaire tout en préservant la qualité audio subjective, même avec des facteurs de compression des données très élevés [45]. Les codecs audio les plus répandus, tels que le codage audio avancé du *Moving Picture Experts Group (MPEG) Advanced Audio Coding (AAC) "MPEG-2/4 AAC"* [46], [47], fournissent un son stéréo de haute qualité à des débits binaires de 96 kbit/s, ce qui correspond à un facteur de réduction des données d'environ 15 fois par rapport à un son de disque compact (CD audio échantillonné à l'origine à 44,1 kHz et 16 bits). Des codecs plus récents (par exemple, *High AAC v2* [48], *Unified Speech and Audio Coding (USAC)* [49] ou *Enhanced Voice Services (EVS)* [50]) offrent une compression encore plus élevée pour une qualité audio comparable. Il est intrigant de reconnaître qu'un codage audio exclusivement basé sur la réduction de la redondance (c'est-à-dire un codage audio sans perte) entraîne une réduction moyenne des données d'environ 2:1 pour le matériel audio de qualité CD. Ainsi, la plus grande partie de la réduction des données fournie par les codecs audio perceptuels peut être attribuée à l'exploitation extensive des propriétés de la perception (réduction de la non-pertinence), pour laquelle les modèles psycho-acoustiques de la perception auditive humaine sont essentiels.

En général, l'approche la plus utilisée du codage audio perceptuel peut être décrite comme suit (Figure 2.9) [45] :

- Le signal audio dans le domaine temporel est transformé en une représentation spectrale sous-échantillonnée à l'aide d'un banc de filtres d'analyse (ou, de manière équivalente, d'une transformée). Ce banc de filtres est généralement échantillonné de manière critique (c'est-à-dire que le nombre d'échantillons de sortie est égal au nombre d'échantillons d'entrée) et (au moins presque) parfaitement reconstruit.

- Un modèle psycho-acoustique (perceptuel) est utilisé pour analyser le signal audio d'entrée et déterminer les aspects perceptuels pertinents du signal, plus particulièrement la capacité de masquage du signal (par exemple, le seuil de masquage) en fonction de la fréquence et du temps. Le résultat est transmis à l'étape de quantification et de codage pour contrôler la distorsion de codage injectée de manière à la rendre inaudible, ou du moins à produire une distorsion audible légère et une gêne minimale. Ce concept de quantification contrôlée perceptivement fait du codeur un codeur audio perceptif.
- Les échantillons spectraux sont ensuite quantifiés et éventuellement codés par entropie pour réduire l'information à une représentation compacte et emballés dans un flux binaire sous forme de valeurs binaires [51].
- Dans le décodeur, le flux binaire est décompressé, le codage entropique est annulé et les valeurs spectrales quantifiées sont ramenées à leur gamme dynamique d'origine et retransformés en un signal de sortie dans le domaine temporel par le banc de filtres de synthèse, c'est-à-dire un banc de filtres qui se comporte de manière complémentaire au banc de filtres d'analyse qui a été utilisé par le codeur.

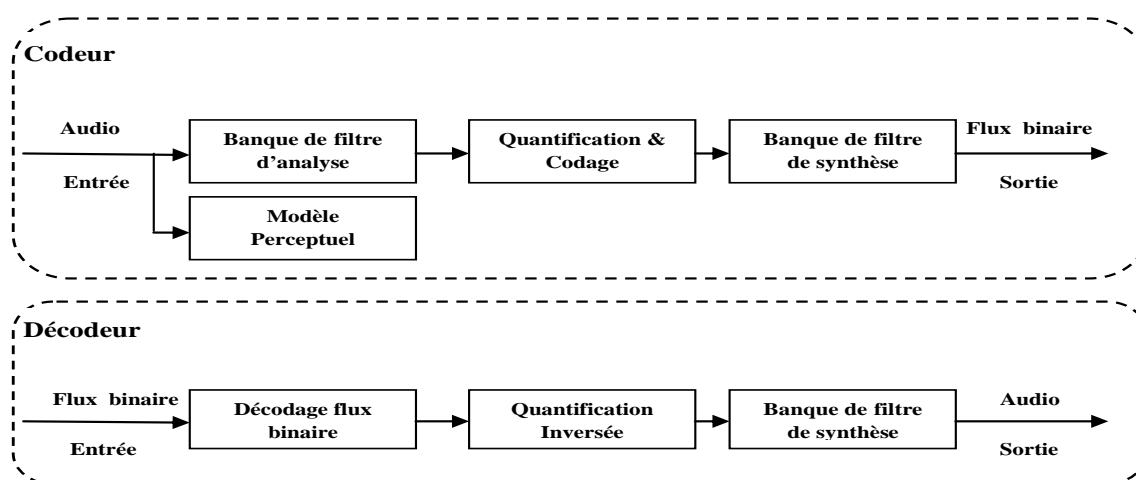


Fig. 2.9 - Schéma de principe d'un Codec audio perceptif

L'idée centrale d'un codec audio perceptif est d'utiliser des critères psycho-acoustiques, tels que des seuils de masquage, pour la quantification des coefficients spectraux afin de maximiser la qualité audio telle qu'elle est perçue par les auditeurs humains, plutôt que d'autres métriques simples d'erreur de type moindres carrés telles que le rapport signal/bruit (SNR), l'erreur quadratique moyenne (MSE – Mean Square Error), et autres. Plus précisément, l'approche du codage audio perceptif peut même impliquer la génération d'un profil de distorsion spectrale de forme perceptive qui fournit une qualité audio subjective améliorée aux dépens du SNR global du signal, par rapport à un codeur qui n'utilise pas de connaissances psycho-acoustiques. En savoir plus sur l'évaluation de la qualité subjective et la dégradation possible de la qualité audio subjective dans les documents suivants BS.1116 [52], Multi-Stimulus Test with Hidden Reference and Anchor (MUSHRA) [53] et "What to listen for" [54].

Les modèles psycho-acoustiques n'ont pas besoin d'être parfaits pour satisfaire leur objectif, et en fait les modèles couramment utilisés ne représentent qu'un petit sous-ensemble des propriétés et des capacités connues du système auditif humain (HAS). D'une part, il semble que plus un modèle perceptif est construit de manière complète et précise par rapport à son homologue naturel (HAS), plus les performances du codage audio perceptif sont élevées. D'autre part, de tels modèles sont souvent

conçus pour fonctionner avec un cadre de codage audio perceptuel de telle sorte que les performances du système de codage global ("encodeur" + "décodeur" = "codec") sont optimisées plutôt que la précision du modèle lui-même. En général, les modèles perceptifs pour le codage audio modélisent les réponses de la HAS au signal audio d'entrée dans un style comportemental (c.-à-d., comme une "boîte noire") plutôt qu'une simulation détaillée des processus physiologiques à l'intérieur de la HAS, en raison de la complexité de calcul prohibitive de tels modèles physiologiques [55].

#### **2.4.4.2 La perception auditive humaine**

Le système auditif humain est assez complexe, mais après avoir effectué un grand nombre de tests psycho-acoustiques, on pense que le système de réponse auditive humaine effectue une analyse de bande critique à court terme et peut être modélisé comme une banque de filtres passe-bande dont les fréquences se chevauchent. Le spectre de puissance n'est pas sur une échelle de fréquence linéaire et les bandes passantes sont de l'ordre de 50-100 Hz pour les signaux inférieurs à 500 Hz et jusqu'à 5000 Hz pour les fréquences plus élevées. Ces bandes de fréquences du système de réponse auditive sont appelées bandes critiques. Jusqu'à 26 bandes critiques couvrant une gamme de fréquences de 24 KHz sont considérées [43].

##### **1) *Le phénomène de masquage***

On observe qu'un signal audio de faible niveau est rendu inaudible, si un signal audio plus fort, dont la fréquence est proche de celle du premier, est présent simultanément. Ce phénomène est connu sous le nom de masquage. Le signal le plus fort qui masque le signal le plus faible est appelé masquant et celui qui est masqué est le masqué. On observe en plus que le masquage est le plus important dans la bande critique dans laquelle se trouve le masquant et, dans une moindre mesure, le masquage est également efficace dans les bandes voisines [42].

Il est possible de définir un seuil de masquage, en dessous duquel la présence d'un son sera rendue inaudible. Le seuil de masquage dépend du niveau de la pression acoustique (SPL), de la fréquence du masquant et des caractéristiques du masquant et du masqué, par exemple si le masquant ou le masqué est un son ou un bruit [42].

Dans la figure 2.10, un signal fort à 1 kHz, agit comme le masquant. Le seuil de masquage (représenté par la ligne continue) diminue considérablement lorsque l'on s'écarte de la fréquence du masquant. Les trois barres pleines ont leurs SPLs inférieurs au seuil de masquage et sont donc masquées. La courbe en pointillés indique le seuil de silence ou d'audition, c'est-à-dire sans la présence d'un masquant. Il est intéressant de noter que le seuil de silence a une valeur plus faible dans la gamme de fréquences (500 Hz - 5 kHz) du spectre audio et que les événements acoustiques dans les zones ombrées ne seront pas audibles.

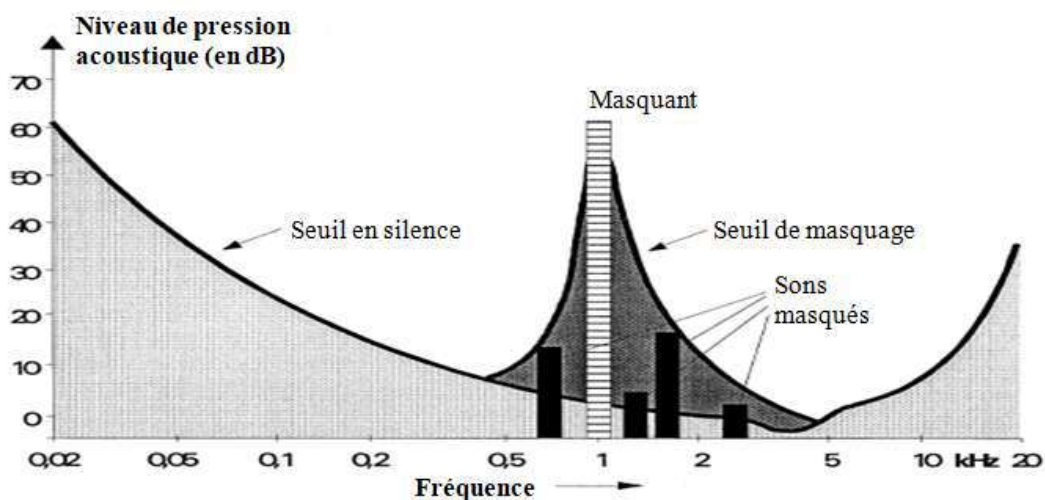


Fig. 2.10 - Seuil de silence et seuil de masquage. [43]

On observe aussi que la pente du seuil de masquage est plus forte vers les basses fréquences, c'est-à-dire que les basses fréquences ne sont pas masquées de la même manière que les hautes fréquences.

Les caractéristiques de masquage sont mesurées par les paramètres suivants [42], [43] :

**a) Rapport signal/masquage (SMR - Signal to mask ratio) :** Le SMR à une fréquence donnée est exprimé comme la différence (en dB) entre le SPL du masquant et le seuil de masquage à cette fréquence.

**b) Rapport masquage/bruit (MNR - Mask to noise ratio) :** Le MNR à une fréquence donnée est exprimé comme la différence (en dB) entre le seuil de masquage à cette fréquence et le niveau de bruit. Pour rendre le bruit inaudible, son niveau doit être inférieur au seuil de masquage, c'est-à-dire le MNR doit être positif.

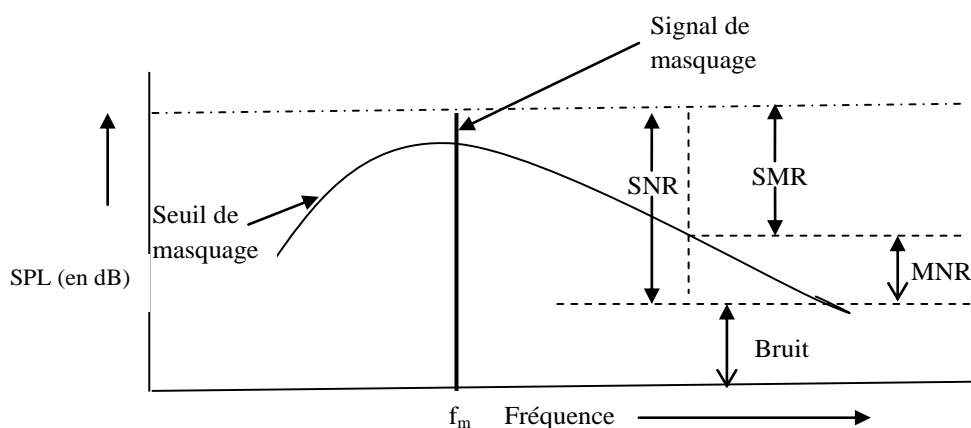


Fig. 2.11 - Caractéristiques du masquage SMR et MNR

Comme l'illustre la figure 2.11, un signal de masquage se produit à la fréquence  $f_m$ , donnant lieu à la courbe du seuil de masquage. Il faut noter que la distance entre le seuil du masquant et le masqué est plus petite dans *bruit-masquant-*



*tonalité* que *tonalité-masquant-bruit* c'est-à-dire que le bruit est un meilleur masquant qu'une tonalité. Dans les codeurs MPEG, les deux seuils jouent un rôle dans le calcul du seuil de masquage. Pour une fréquence  $f$  donnée, on indique le SMR, le rapport signal/bruit (SNR) et le MNR correspondant au niveau de bruit. Il est évident que :

$$\text{SMR}(f) = \text{SNR}(f) - \text{MNR}(f)$$

Nous avons jusqu'à présent considéré un seul masquant. S'il y a plus d'un masquant, alors chaque masquant contribue à son propre seuil de masquage et un seuil de masquage global est calculé qui décrit la distorsion juste perceptible en fonction de la fréquence.

## 2) Masquage temporel

Le phénomène de masquage décrit dans la sous-section précédente est également appelé masquage simultané, où l'on suppose que le masquant et le masqué apparaissent au même instant. Le masquage est également observé lorsque deux sons apparaissent dans un court intervalle de temps et que le plus fort masque le plus faible. Ce phénomène est connu sous le nom de masquage temporel. Outre le masquage simultané, le masquage temporel joue également un rôle majeur dans la perception auditive humaine [42].

Le masquage temporel est possible, même si le masqué précède le masquant d'un court intervalle de temps et est associé au pré masquage et au post masquage où le premier a une durée d'un dixième par rapport au second. L'ordre de durée du post masquage est de 50 à 200 ms [42] (Figure 2.12).

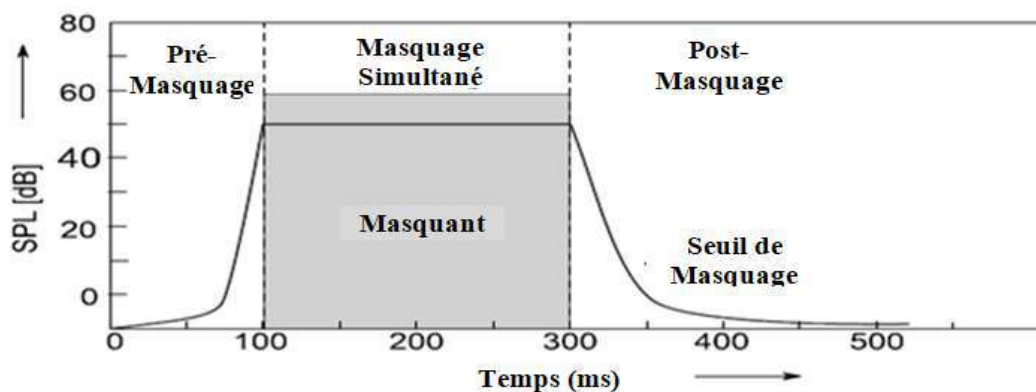


Fig. 2.12 - Illustration des effets du masquage temporel [42]

### 2.4.4.3 Codage perceptuel dans le MPEG audio

Un algorithme de codage de source audio efficace doit remplir les deux objectifs de base suivants [42]:

- la suppression de la redondance, dans laquelle les redondances statistiques entre les échantillons adjacents sont exploitées, et
- la suppression de la non-pertinence, qui est motivée par la perception, puisque tout ce que nos oreilles ne peuvent pas entendre peut être supprimé.

Dans la suppression de la non-pertinence, les phénomènes de masquage simultané et temporel jouent un rôle dominant dans le codage audio MPEG. Nous avons déjà

noté que le niveau de bruit doit être inférieur au seuil de masquage. Comme le bruit dû à la quantification est fonction du nombre de bits auxquels les échantillons sont quantifiés, l'algorithme d'allocation des bits doit tenir compte des caractéristiques de masquage. La figure 2.13 montre le schéma fonctionnel d'un codeur basé sur la perception, qui exploite le phénomène de masquage [42].

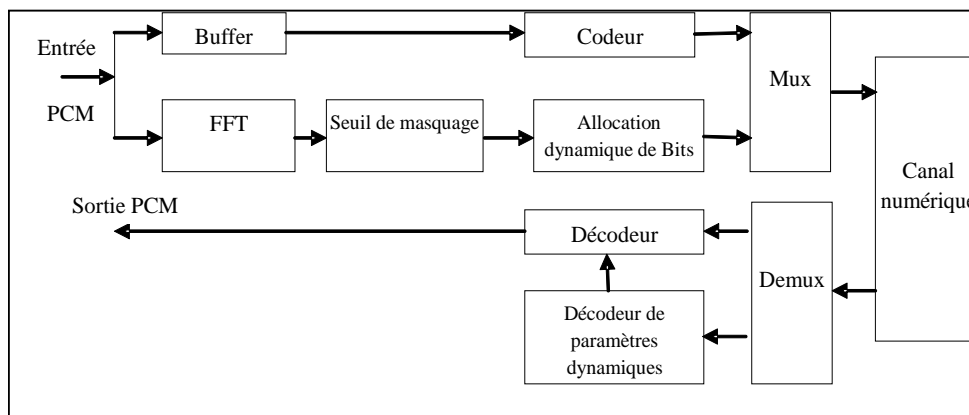


Fig. 2.13 - Schéma de principe des codeurs basés sur la perception (Structure of MPEG-1 audio encoder and decoder -Layers I et II)

Comme indiqué dans la figure 2.13, une transformation de Fourier rapide (FFT) est calculée à partir des échantillons audio PCM entrants et le spectre audio complet est obtenu, à partir duquel les composantes sonores des signaux de masquage peuvent être déterminées. A partir de là, un seuil de masquage global ainsi que le SMR dans l'ensemble du spectre audio sont calculés. L'allocateur dynamique de bits pour coder le flux binaire utilise l'information SMR. Il est possible de rendre le schéma de codage transparent sur le plan perceptif en maintenant le bruit de quantification en dessous du seuil de masquage global. Le décodage s'effectue en sens inverse de celui d'encodage, mais sans passer par le modèle psycho acoustique.

Il convient toutefois de noter que les connaissances en matière de calcul du seuil de masquage global sont limitées dans le sens où le modèle perceptif ne prend en compte que les masques simples et stationnaires et qu'il peut échouer dans certaines situations pratiques. La solution à ce problème est de garder une marge de sécurité suffisante.

## 2.5 CONCLUSION

Dans ce chapitre, nous avons présenté le son, ses caractéristiques physiques, sa numérisation, sa compression en tant que signal audio et les différents formats des fichiers audio non compressés, compressés sans perte et compressés avec perte. Nous avons montré certaines des idées de base relatives au codage audio. Les concepts du codage audio perceptuel permettent de réaliser des réductions considérables de débit binaire dans la représentation de matériel audio de haute qualité. Une grande partie de la réduction du débit binaire peut être attribuée à l'exploitation des effets psycho-acoustiques et aux méthodes de traitement du signal appropriées, notamment [45] :

- La mise en forme du bruit de l'erreur de quantification en fréquence et en temps pour obtenir le meilleur masquage possible du signal d'erreur, et donc une qualité subjective optimale

- Le contrôle de l'utilisation d'outils de codage optionnels pour améliorer les performances du codec.

Dans un codec audio, le modèle psycho acoustique est d'une importance capitale pour la mise en œuvre de ces fonctions. Les aspects perceptifs du masquage ont été utilisés particulièrement dans la norme MPEG. Nous avons choisi de parler du codeur MPEG car c'est un codeur normalisé et le plus utilisé dans le monde de l'audio numérique.

Les codeurs MPEG/Audio sont contrôlés par des modèles psycho acoustiques qui peuvent être améliorés, laissant ainsi la place à une amélioration évolutive des codecs. À l'avenir, nous verrons de nouvelles solutions pour le codage. Une meilleure compréhension de la perception auditive humaine et de la présentation stéréo conduira à de nouvelles propositions.

L'audio numérique multi canal améliore les images stéréophoniques et sera important à la fois pour les applications audio et les autres applications multimédias. Les recherches en cours permettront d'améliorer les représentations multicanaux en utilisant mieux les corrélations entre les canaux et les effets de masquage pour réduire encore les débits binaires au plus bas.

Nous pouvons également nous attendre à des solutions pour des présentations spéciales destinées aux personnes souffrant de déficiences auditives ou visuelles, qui pourront utiliser les configurations multicanaux de diverses manières.

Les activités émergentes du groupe d'experts de l'ISO/MPEG visent à proposer un codage audio qui offrira des taux de compression plus élevés et qui intégrera toute la gamme des sons, du codage audio haute fidélité au codage de la parole, en passant par la parole synthétique et les sons synthétiques (ISO/IEC MPEG-4) [56]. Comme la qualité audio de base est plus importante que la compatibilité avec les normes existantes ou à venir, cette activité ouvrira la voie à des solutions totalement nouvelles.

Historiquement, de nombreux progrès ont été réalisés au fil des années tant dans le codage audio que dans ses modèles perceptuels. Avec la disponibilité de plus de ressources informatiques, les nouvelles générations de modèles perceptuels [45] ont évolué, incluant de nouveaux aspects tels que la perception de la modulation, et sont sur le point de prouver leur utilité.

Les futures générations de modèles perceptifs pourraient être confrontées à de nouveaux défis dans le contexte de l'audio comprimé pour les applications de réalité virtuelle (VR) à 6 degrés de liberté (6DoF) [57], [58], où, au-delà de l'audio spatial, la représentation visuelle et le mouvement de l'utilisateur deviennent très pertinents et doivent être compris et intégrés dans le cadre d'une expérience multimodale. En attendant l'évolution de la compréhension des aspects perceptifs pertinents du codage, et du traitement audio dans de tels scénarios perceptifs hautement immersifs [45].

### 3. Etat de l'Art sur les Techniques du Tatouage Audio Numérique

---

#### Sommaire

---

3.1	<b>Introduction</b> .....	39
3.2	<b>Concept du tatouage audio numérique</b> .....	39
3.3	<b>Classification des systèmes de tatouage audio numérique</b> .....	45
3.4	<b>Les techniques fondamentales de tatouage audio numérique</b> .....	51
3.5	<b>Travaux connexes dans le domaine de tatouage audio numérique</b> .....	75
3.6	<b>Conclusion</b> .....	78

---

## 3.1 INTRODUCTION

L'objectif de ce chapitre est de présenter les aspects principaux et les techniques liées aux évolutions de la technologie de tatouage audio numérique traitées dans le contexte de cette thèse. Après avoir donné un aperçu sur les techniques de dissimulation de l'information dans le premier chapitre introduction générale, nous présenterons le tatouage numérique audio et ses différentes étapes qui conduisent à l'insertion de la marque et sa détection/extraction. Nous décrirons également les caractéristiques et les propriétés requises pour un système de tatouage audio, et la classification de ces systèmes selon différents critères. Ensuite nous présenterons les principales techniques de tatouage audio dans les différents domaines d'insertion. A la fin de ce chapitre nous nous passerons en revue les différents travaux connexes que nous avons utilisés pour la comparaison de nos approches réalisées dans le cadre de cette thèse.

## 3.2 CONCEPT DU TATOUAGE AUDIO NUMERIQUE

### 3.2.1 Concepts de base et termes généraux

Le tatouage numérique simple est une technologie dans laquelle une signature ou une marque numérique (*digital watermark*) est insérée en général d'une manière invisible dans un signal original dit aussi signal hôte (*host signal*). Cette marque contient des données qui peuvent être employées dans diverses applications comme la protection de copyright, la surveillance d'émission, l'authentification de données ou la transmission sécurisée. La marque numérique (qui est essentiellement une séquence aléatoire de bruit, un logo binaire, une image à niveaux de gris, ou un message de chaînes de caractères) est insérée dans le signant original via l'*encodeur* (*watermark embedder*). La marque qui est perceptuellement transparente, son existence n'est indiquée que lorsque le média tatoué est passé par un *détecteur* approprié (*watermark detector*). Un encodeur a deux entrées ; l'une est la marque numérique (généralement accompagnée d'une clé secrète) et l'autre est le signal original (qui est un fichier audio dans notre cas). La sortie de l'encodeur est le signal tatoué (*Watermarked signal*) qui ne doit pas être perceptuellement différent du signal original. Le signal tatoué est enregistré ou diffusé et plus tard présenté au détecteur qui déterminera si la marque est présente dans le fichier audio tatoué, et si oui elle sera extraite. La figure 3.5 montre le schéma général d'un tatouage audio numérique [9, p. 4].

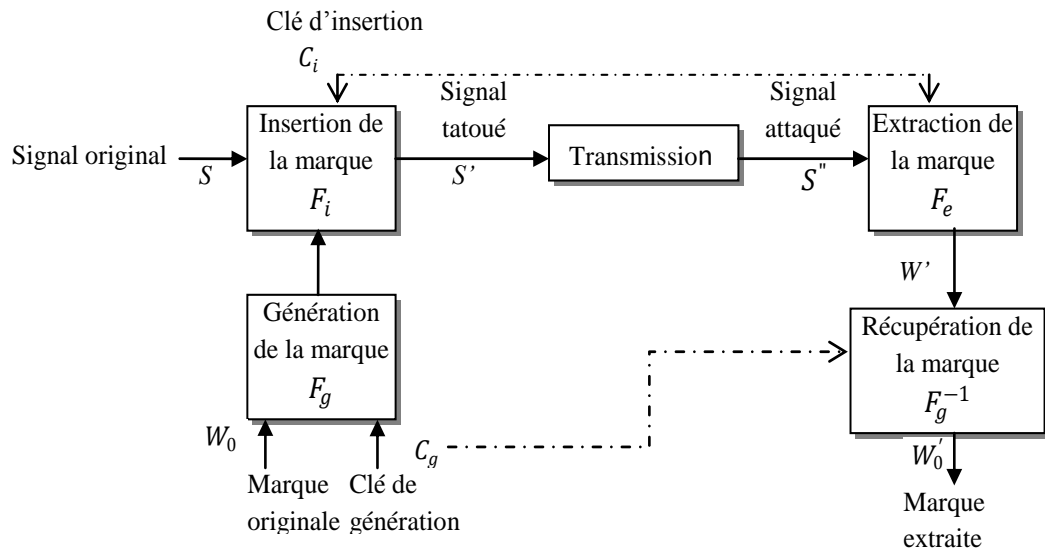


Fig. 3.1 - Schéma général d'un tatouage audio numérique

### 3.2.2 Schéma général du tatouage numérique

La technique de tatouage numérique simple se compose de deux processus : un processus d'insertion de la marque et un processus de détection ou d'extraction de cette marque comme illustré dans la figure 3.1.

#### 3.2.2.1 Processus d'insertion

L'insertion de la marque originale  $W_0$  dans le signal audio original  $S$  permet d'obtenir un signal audio tatoué noté  $S'$ . Mais généralement dans la phase d'insertion les systèmes de tatouage utilisent une ou plusieurs clés cryptographiques pour assurer une sécurité contre l'insertion ou l'extraction non autorisée de la marque. Généralement ce processus peut être divisé en deux processus : la génération de la marque et son insertion.

##### 1) La génération de la marque

La marque numérique à insérer  $W$  est générée à partir d'une marque initiale  $W_0$  et d'une clé de génération  $C_g$ , en utilisant une fonction de génération  $F_g$ , généralement cette fonction  $F_g$  est une fonction cryptographique. La marque originale peut prendre plusieurs formes, elle peut être une image binaire ou en niveaux de gris, une séquence aléatoire de bruit, ou une suite de caractères. La fonction de génération s'exprime comme suit :

$$W = F_g(W_0, C_g) \quad (3.1)$$

La marque peut aussi dépendre du signal  $S$  dans lequel elle est insérée

$$W = F_g(W_0, C_g, S) \quad (3.2)$$

2) **L'insertion de la marque :**

Le signal audio tatoué  $S'$  est obtenu par l'insertion de la marque générée  $W$  dans le signal original  $S$  par une fonction d'insertion  $F_i$ , la fonction d'insertion s'exprime comme suit :

$$S' = F_i(\mathfrak{X}(S), W) \quad (3.3)$$

Une clé cryptographie  $C_i$  peut aussi être utilisée :

$$S' = F_i(\mathfrak{X}(S), W, C_i) \quad (3.4)$$

L'espace d'insertion  $\mathfrak{X}$  peut être le domaine temporel ou le résultat d'une transformation réversible fréquentielle ou Cepstral. La fonction d'insertion de la marque  $F_i$  peut être une fonction multiplicative, additive ou une substitution. Une fonction multiplicative ou additive insère la marque dans le signal original en la multipliant ou en l'ajoutant à ce dernier, en revanche dans une fonction de substitution, la marque à insérer est substituée à des composants du signal audio original. La figure 3.2 montre le processus d'insertion d'une marque.

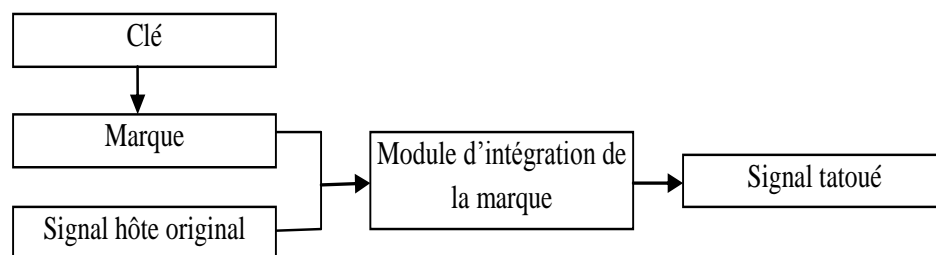


Fig. 3.2 - Processus d'insertion d'une marque

**3.2.2.2 Processus d'extraction**

Le processus de détection/extraction prend en entrée le signal audio tatoué et éventuellement attaqué  $S''$  et les clés  $C_i$  et  $C_g$ . Selon l'algorithme d'extraction le signal audio original peut être ou non nécessaire lors de la détection. Si le signal audio original n'est pas présent la détection est dite aveugle ou non-informée

$$W' = F_e(S'', C_i) \quad (3.5)$$

Dans le cas contraire, la détection est dite non aveugle ou informée

$$W' = F_e(S'', S, C_i) \quad (3.6)$$

La marque originale est finalement récupérée à partir de la marque  $W'$  et de la clé  $C_g$  en utilisant la fonction de génération inversée  $F_g$

$$W_0 = F_g^{-1}(W', C_g) \quad (3.7)$$

Généralement la robustesse des schémas non aveugles est plus importante que les schémas aveugles. Le signal audio original fournit une référence pouvant améliorer le processus de détection.

La figure 3.3 montre le processus de détection/extraction d'une marque en utilisant un schéma de tatouage non aveugle.

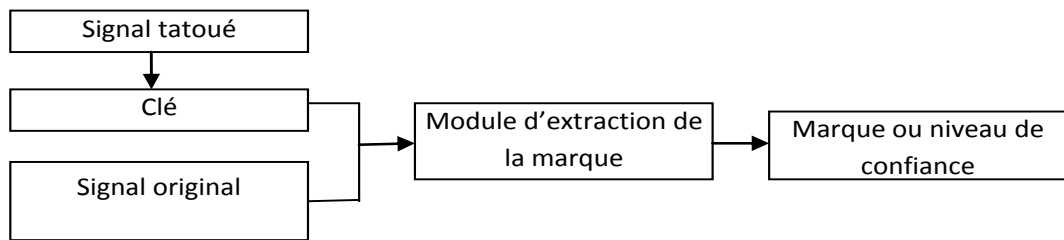


Fig. 3.3 - Processus de détection/extraction d'une marque

### 3.2.2.3 Les types de marque et leur détection

Il existe trois principaux types de marques qui peuvent être intégrées dans un signal audio [59] :

#### 1) Séquence gaussienne pseudo-aléatoire

Une marque de séquence gaussienne est une séquence de nombres comprenant 1 et -1 et ayant un nombre égal des 1's et des -1's. Elle est appelée filigrane avec une moyenne de zéro et une variation de un. De tels filigranes ou marques sont utilisés pour une détection objective en utilisant une mesure de corrélation.

#### 2) Marques d'images binaires ou d'images en niveaux de gris

Certains algorithmes de tatouage intègrent des données significatives sous la forme d'une image logo au lieu d'une séquence gaussienne pseudo-aléatoire. De telles marques sont appelées marques d'image binaire ou marques d'image en niveaux de gris. Ces marques sont utilisées pour la détection subjective.

En fonction du type de la marque insérée, un décodeur approprié doit être conçu pour détecter la présence de la marque. S'il s'agit d'une hypothèse de séquence gaussienne pseudo-aléatoire, un test est effectué pour détecter la présence de la marque. Supposons que  $W$  est la séquence binaire originale de la marque et  $W'$  est la séquence binaire extraite de la marque, nous pouvons alors calculer le taux d'erreur sur les bits (BER) pour détecter la présence de la marque. Si le BER est égal à zéro, il indique la présence de la marque, en revanche, s'il est égal à un, il indique l'absence de la marque. BER est calculé comme suit. Supposons que  $D$  est le signal de la marque récupéré et  $N$  le nombre de bits de la marque, alors [59]:

$$D = \begin{cases} 1 & \text{si } W_i \neq W'_i \\ 0 & \text{si } W_i = W'_i \end{cases} \quad \text{BER}(W, W') = \frac{\sum D}{N} \quad (3.8)$$

Le coefficient de corrélation normalisée peut aussi être utilisé pour savoir le taux de similarité entre la marque extraite et la marque originale

$$\text{NC}(W, W') = \frac{\sum W W'}{\sqrt{\sum W_i^2} \sqrt{\sum W'_i^2}} \quad (3.9)$$



### **3.2.3 Les propriétés requises d'un système de tatouage audio numérique**

Les propriétés les plus importantes des techniques de tatouage numérique audio sont la transparence ou l'invisibilité, la robustesse, la capacité, la sécurité, la fiabilité de la détection et la complexité. La transparence se rapporte aux propriétés du sens humain, dans le cas du tatouage audio, c'est le système auditif humain HAS qui est pris en considération. Un tatouage transparent ne doit causer aucune distorsion ni perte de qualité [60].

L'optimisation de ces propriétés est mutuellement concurrentielle et ne peut être réalisée en même temps. Si nous voulons intégrer un grand message, nous ne pouvons pas exiger une grande robustesse en même temps. Un compromis raisonnable est toujours nécessaire. D'un autre côté, si la robustesse à forte distorsion est un problème, le message qui peut être caché de manière fiable ne doit pas être trop long.

La Fédération internationale de l'industrie phonographique (IFPI- *International Federation of Phonographic Industry*) déclare qu'un système de tatouage audio numérique efficace doit satisfaire aux exigences suivantes, appelées propriétés des systèmes de tatouage audio numérique [61].

#### **3.2.3.1 Imperceptibilité**

L'algorithme de tatouage doit insérer la marque sans changer la qualité perceptive du signal original. La marque ajoutée ne doit pas dégrader la qualité du contenu. Dans le cas du tatouage audio numérique, cette propriété est liée à l'inaudibilité c'est-à-dire l'information tatouée ne doit pas dégrader perceptuellement le signal audio dans lequel elle est insérée, son insertion doit être transparente.

L'imperceptibilité est évaluée à l'aide des mesures mathématiques objectives comme le rapport signal sur bruit (Signal noise rate –SNR) ou subjectives comme les tests d'écoute. Le SNR mesure la sensibilité du signal tatoué par rapport au bruit du fond. Pour tout algorithme de tatouage audio, l'IFPI impose un SNR supérieur à 20 dB [61].

Cependant, si un signal est vraiment imperceptible, les algorithmes de compression avec perte basés sur la perception introduisent soit d'autres modifications qui dépassent le seuil d'audition ou suppriment un tel signal, Il est alors important de développer des techniques qui peuvent être utilisées pour ajouter des signaux de marques imperceptibles ou non détectables dans des zones perceptivement significatives pour contrer les effets de traitement du signal [60].

#### **3.2.3.2 Robustesse**

La robustesse se réfère à ce que la marque insérée dans le signal audio original a la capacité de survivre après une variété d'opérations de traitement et d'attaques. Donc le tatouage doit être robuste pour le traitement général des signaux, la transformation géométrique, le filtrage, la compression, l'ajout de bruit et les attaques malveillantes. Selon l'application, la technique du tatouage peut supporter différents niveaux de robustesse contre les changements sur le signal tatoué causés par des attaques bienveillantes ou malveillantes. Si le tatouage est dédié à la

protection de droit d'auteur, alors la marque doit être robuste contre toute manipulation, car elle permet dans ce cas d'attester la propriété du document numérique. Cependant, si le tatouage vise l'authentification du contenu, la robustesse de la marque doit être contrôlée : la marque peut tolérer certaines manipulations du document numérique, mais doit être insensible aux traitements classiques comme le filtrage et la compression [60].

Il est important de noter que la robustesse n'inclut pas les attaques sur le schéma d'insertion qui sont basées sur la connaissance de l'algorithme du tatouage ou sur la disponibilité de la fonction de détection (la clé secrète). La robustesse signifie la résistance aux modifications "aveugles" et non ciblées, ou aux opérations courantes sur le support. Par exemple, les outils *Stirmark* ou *Mosaik* attaquent la robustesse des algorithmes de tatouage par des distorsions géométriques. Pour la reconnaissance des manipulations, le tatouage doit être fragile pour détecter les supports altérés [60].

### 3.2.3.3 Capacité

La capacité indique le nombre de bits d'information pouvant être insérés. Elle montre également la possibilité d'insérer plusieurs marques dans un document en parallèle. La capacité est mesurée en nombre de bits insérés par unité de temps. Généralement elle est donnée en bits par seconde (*bps*). Cette quantité de bits à cacher varie selon l'application du tatouage : de quelques bits pour cacher un identifiant dans une application de protection de droit d'auteur, à des milliers de bits pour une application de transmission de données [60].

La condition de capacité se heurte toujours à deux autres conditions importantes, c'est-à-dire l'imperceptibilité et la robustesse (Figure 3.4) [62]. Une capacité plus élevée est généralement obtenue au détriment de la robustesse ou de l'imperceptibilité, ou les deux. Il faut nécessairement trouver un compromis entre ces trois paramètres en fonction de l'application envisagée.

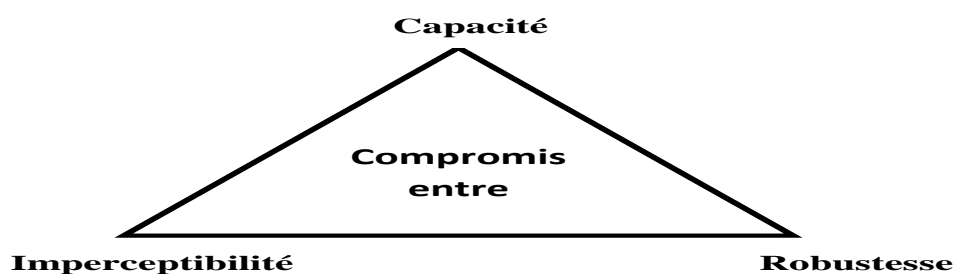


Fig. 3.4 - Compromis entre robustesse, capacité et imperceptibilité

### 3.2.3.4 Sécurité

La sécurité décrit si les informations de marque insérées ne peuvent pas être supprimées au-delà de la détection fiable par des attaques ciblées basées sur une connaissance complète de l'algorithme d'insertion et de détection, à l'exception de la clé, et la connaissance au moins d'une des données tatouées. La sécurité du tatouage exige que la marque doive être difficile à enlever ou à modifier sans endommager le signal hôte. Comme tous les systèmes de tatouage cherchent à protéger l'information sur les marques, sans perte de qualité, la sécurité du tatouage peut être considérée

comme la capacité d'assurer la confidentialité et l'intégrité de l'information sur les marques, et de résister aux attaques malicieuses [63].

### 3.2.3.5 Complexité et Coût de calcul

Afin de réduire le coût de calcul, une méthode de tatouage devrait être moins complexe. Les méthodes de tatouage avec des algorithmes très complexes nécessiteront plus de logiciels ainsi que des ressources matérielles et entraîneront donc un coût de calcul plus élevé. La simplicité de calcul est généralement préférée dans les environnements contraints en ressources comme les systèmes mobiles et aussi pour les applications en temps réel. Par exemple dans le contrôle de diffusion, l'insertion et la détection doivent être faites en temps réel, tandis que dans les applications de protection de droit d'auteur le temps n'est pas un facteur crucial pour l'exécution pratique [60].

### 3.2.3.6 Détection aveugle ou non aveugle

Une technique de tatouage est dite aveugle, si elle ne nécessite pas le signal audio original pour récupérer la marque du signal tatoué. Inversement, une technique de tatouage est dite non aveugle, si elle a besoin du signal hôte original pour extraire la marque du média tatoué. Les systèmes de tatouage non aveugle sont plus robustes que les systèmes de tatouage aveugle en raison de la disponibilité du signal hôte original au moment de la détection. Cependant, les systèmes de tatouage aveugle sont plus souples et plus pratiques, ils réduisent le coût de calcul et la mémoire pour le stockage du signal audio original [60].

## 3.3 CLASSIFICATION DES SYSTEMES DE TATOUAGE AUDIO NUMERIQUE

Nous pouvons classer le tatouage audio numérique en cinq catégories sur la base de divers critères [3] comme montre la figure 3.5 :

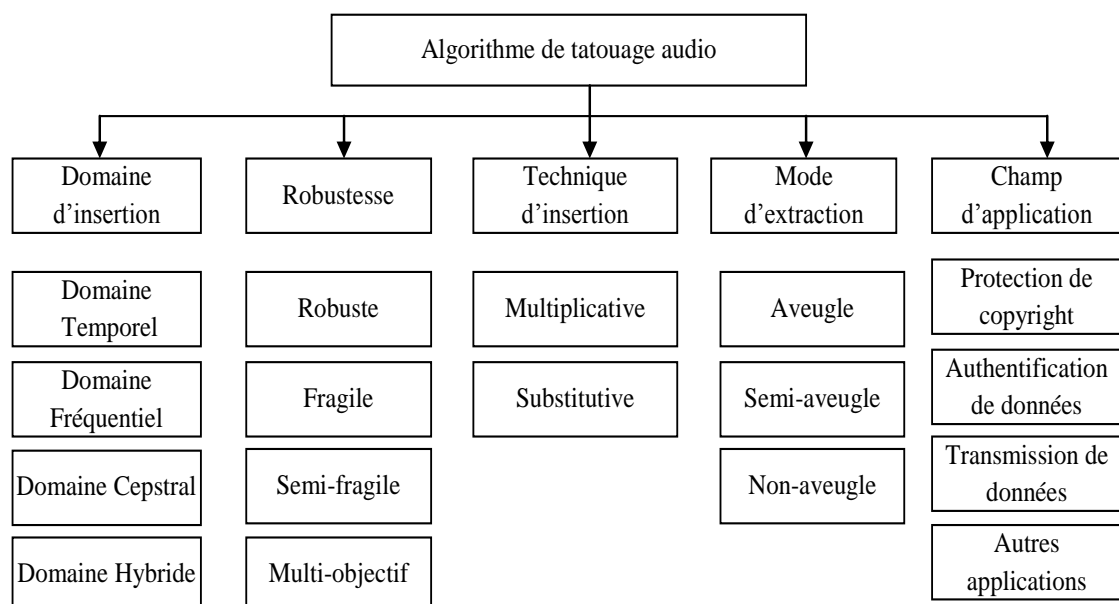


Fig. 3.5 - Classification de tatouage audio numérique

### 3.3.1 Selon le domaine de travail

Une première façon de classer le tatouage est par la manière dont la marque est insérée dans le fichier audio de couverture, elle peut être insérée soit dans le domaine spatial (appelé aussi temporel), dans le domaine de transformation (appelé aussi fréquentiel) ou dans le domaine Cepstral. Cependant, il existe certains schémas qui combinent deux ou même trois domaines, on peut les classer dans un autre domaine : le domaine hybride.

#### 3.3.1.1 Insertion dans le domaine spatial

Cette technique de tatouage se focalise sur la modification directe des valeurs des échantillons du signal audio hôte qui sont sélectionnés aléatoirement. Elle charge directement les données brutes de la marque dans les valeurs de l'échantillon. Certains de ses algorithmes sont *LSB (Least Significant bit)*, *SSM* technique d'étalement du spectre basée sur la modulation, tatouage additif etc. Ces algorithmes sont simples et faciles à implémenter et peu coûteux en temps de calcul. Ils peuvent être utilisés pour les applications en temps réel (authentification). Par contre la quantité d'information à insérer doit être faible pour assurer l'imperceptibilité de la marque. En plus, la robustesse des algorithmes de tatouage numérique conçus dans ce domaine est limitée, ils ne montrent pas assez de robustesse aux altérations courantes du signal [61].

#### 3.3.1.2 Insertion dans le domaine de transformation ou fréquentiel

La deuxième technique est le domaine de transformation dans lequel d'abord le signal original est transformé en domaine de fréquence par l'utilisation de n'importe quelle méthode comme une transformée de Fourier Discrète (*DFT*), une transformée en Cosinus Discrète (*DCT*), une transformée en Ondelette Discrète (*DWT*), ou une transformée hybride. La marque est ensuite insérée en modifiant les coefficients d'une transformée fréquentielle du signal hôte. Enfin, la transformation inverse correspondante est faite pour obtenir le signal tatoué. Ces schémas sont plus robustes et plus complexes mais largement utilisées. Ils peuvent être utilisés pour la protection des droits d'auteur. Ces transformées peuvent être utilisées avec des techniques d'insertion du domaine spatial comme *LSB*, étalement du spectre, la quantification de l'indice de Modulation (*QIM*) et *Patchwork* [61].

La comparaison dans le tableau 3.1 montre que le principal avantage d'insérer la marque dans le domaine de transformation est que la technique basée sur ce domaine est plus robuste par rapport au domaine spatial [62].

Tableau 3.1 - Comparaison entre domaine spatial et domaine fréquentiel

<i>Critère</i>	<i>Domaine spatial</i>	<i>Domaine de transformation</i>
Cout de calcul	Faible	Elevé
Robustesse	Fragile	Plus robuste
Qualité de la perception	Contrôle élevé	Contrôle faible
Complexité de calcul	Faible	Elevé
Temps de calcul	Réduit	En plus
Capacité	Elevée (dépend de la taille du fichier hôte)	Faible
Exemple d'application	Authentification	Droits d'auteur

### 3.3.1.3 Insertion dans le domaine Cepstral

La transformée de Cepstrum consiste en trois étapes linéaires consécutives illustrées à la figure 3.6. Ces étapes sont [64] :

1. la transformée de Fourier rapide
2. le logarithme
3. la transformée de Fourier rapide inverse

Toutes ces étapes peuvent être inversées et le même signal dans le domaine temporel peut être récupéré par la transformée de Cepstrum inverse, comme le montre la figure 3.7.

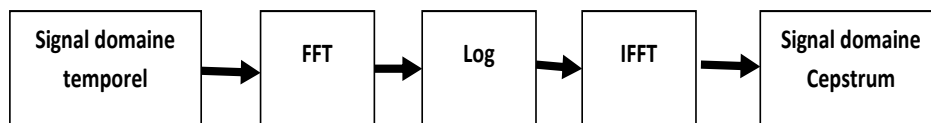


Fig. 3.6 - Transformation du domaine de Cepstrum

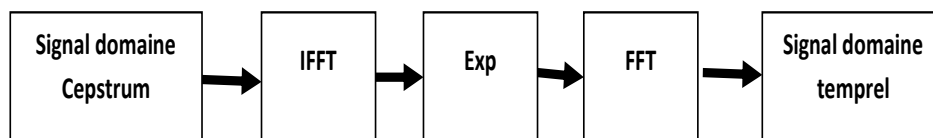


Fig. 3.7 - Transformation inverse du domaine de Cepstrum

La transformée de Cepstrum est largement utilisée dans le traitement et la reconnaissance de la parole [64]. Des études expérimentales ont montré que le domaine du cepstral présente une variance moindre par rapport à la plupart des traitements de signaux courants. Cette caractéristique d'invariance aux attaques du domaine cepstral le place au premier rang des domaines utilisés dans les systèmes de tatouage audio.

### 3.3.1.4 Algorithmes hybrides

Les performances des méthodes insérant la marque directement dans le domaine temporel ou en utilisant une transformée fréquentielle sont souvent limitées. Ainsi des algorithmes hybrides sont proposés. Le terme hybride peut signifier : (1) la combinaison de deux domaines différents, (2) la combinaison de plusieurs transformées fréquentielles, (3) la combinaison de l'un des trois domaines avec une transformation matricielle.

## 3.3.2 Selon la perception humaine

Selon la perception humaine, les algorithmes de tatouage audio peuvent être classés en trois catégories : robuste, fragile et semi-fragile [9]. Néanmoins il ya des algorithmes qui assurent les trois objectifs à la fois, ils sont classés dans la catégorie multi objectif :

### **3.3.2.1 Tatouage robuste**

La plupart des algorithmes de tatouage audio qu'on trouve dans la littérature font partie de la catégorie robuste, qu'ils sont destinés à la protection des droits d'auteurs ou à la transmission de données. Ils font l'objet de plusieurs recherches afin de concevoir le schéma le plus robuste possible. En tatouage inaudible, la marque est insérée de telle manière qu'elle ne peut pas être perçue par l'oreille humaine. Elle est utilisée pour protéger le fichier audio contre la violation des droits d'auteur et empêcher la copie non autorisée. La marque insérée peut être extraite ou détectée plus tard pour identifier le propriétaire ou l'origine du fichier audio hôte, donnant ainsi des informations pertinentes. La présence d'une marque inaudible ne doit pas nuire à la protection du fichier tatoué.

La marque insérée peut résister au traitement du signal, et la compression avec perte. Elle n'est pas détruite après une attaque et peut toujours être détectée pour fournir la preuve de propriété. Cela signifie que la marque doit être insérée dans les composants les plus perceptivement significatifs du fichier audio.

### **3.3.2.2 Tatouage fragile**

En tatouage fragile si le fichier est modifié, alors la marque est facilement détruite. Le tatouage fragile est principalement utilisé pour la protection de l'intégrité et pour l'authentification, qui doit être très sensible aux changements du signal. Nous pouvons déterminer si les données ont été altérées en fonction de l'état de tatouage fragile.

### **3.3.2.3 Tatouage semi-fragile**

On trouve aussi un type de tatouage considéré comme semi-fragile, capable de tolérer un certain degré de changement à un signal tatoué, comme l'ajout d'un bruit de quantification suite à une compression avec perte.

## **3.3.3 Selon la technique d'insertion**

Les algorithmes de tatouage audio peuvent être aussi classés en deux catégories selon la technique d'insertion : algorithme multiplicatif (ou additif) et algorithme substitutif. Dans la première catégorie, l'insertion se fait par une multiplication (ou addition) entre la marque et le signal original, tandis que dans l'insertion substitutive, la marque est insérée en substituant les composantes du signal hôte.

### **3.3.3.1 Systèmes de tatouage additifs**

Dans les systèmes de tatouage additifs le message est ajouté au support audio. Ceci peut être réalisé dans le domaine temporel ou fréquentiel, ou bien dans d'autres domaines moins utilisés tels que le domaine cepstral. La plupart de techniques existantes portent sur les deux premiers domaines.

### **3.3.3.2 Systèmes de tatouage substitutifs**

Les techniques utilisées dans ce type de système consistent à substituer quelques composants du son par la marque désirée. Plusieurs critères sont envisageables selon

le type de schéma utilisé pour représenter le signal. L'idée peut consister à chercher la méthode qui s'adapte le mieux aux conditions d'imperceptibilité et/ou de robustesse.

### **3.3.4 Selon le champ d'application**

On peut distinguer les algorithmes de tatouage selon leur champ d'application. Dans le tatouage audio, les trois applications les plus répandues sont la protection des droits d'auteur, l'authentification de données et la transmission de données particulièrement dans le domaine de la télémédecine. Le niveau de robustesse exigée pour ces différentes applications est variable [62].

#### **3.3.4.1 Protection des droits d'auteur**

C'est l'une des premières applications et la plus importante du tatouage numérique. Le contenu numérique peut être intégré avec une marque contenant des métadonnées identifiant les détenteurs de droits d'auteur, afin d'empêcher d'autres parties, en cas de litige, de réclamer sa propriété. Cette application exige un niveau élevé de robustesse pour s'assurer que la marque insérée ne peut être enlevée sans causer une déformation significative dans le signal audio numérique. Ce service est toujours d'actualité et concerne encore la majorité de publication des contenus numériques.

#### **3.3.4.2 Authentification et contrôle d'intégrité**

Pour la problématique d'authentification et de contrôle d'intégrité, on insère dans le fichier audio une information qui certifie que ce dernier n'a pas été modifié. Il s'agit d'un tatouage fragile ou semi fragile pour tolérer certaines modifications du signal audio tel que les processus de traitement de signal.

#### **3.3.4.3 Transmission de données**

Le tatouage audio peut être utilisé pour une transmission de données sécurisées. Dans ce genre d'application, l'information insérée n'est pas nécessairement liée au signal audio. Contrairement au tatouage pour la protection des droits d'auteur qui doit être robuste, le tatouage de transmission de données doit garantir une haute capacité pour pouvoir véhiculer un maximum de données. La contrainte de robustesse est par conséquent moins forte.

#### **3.3.4.4 Applications médicales**

Le tatouage audio et image peut être appliqué pour protéger le droit d'auteur des images médicales et des fichiers audio médicaux comme les sons respiratoires et les battements de cœur. Les informations peuvent être protégées de l'accès illégal par des techniques de tatouage. Ces applications incluent l'imagerie médicale, la télésanté et la télémédecine, entre autres. L'imagerie médicale permet de visualiser les tissus, les organes ou d'autres parties du corps, en utilisant les technologies de l'information et des communications.

La télésanté comprend la télé chirurgie, le télédiagnostic, les téléconférences et d'autres applications médicales. La télémédecine met en relation des spécialistes et

des patients séparés géographiquement. Par conséquent, pour garantir la confidentialité, l'authenticité, l'intégrité et la disponibilité associées à l'échange de données du dossier médical électronique du patient, des techniques de tatouage appropriées peuvent être utilisées.

Le message intégré a une grande importance et l'objet hôte (c'est-à-dire l'image médicale ou le son audio de battements de cœur ou respiratoire) est également important. L'objet hôte doit conserver sa qualité sans déformation, car sa modification peut avoir un impact négatif sur la vie du patient en provoquant des erreurs de diagnostic et de traitement. Par conséquent, des systèmes de tatouage réversibles [65] ont été créés pour remédier à ce problème en intégrant une technologie capable de restaurer à la fois la marque interne et le fichier hôte original.

La marque peut être utilisée pour écrire le nom unique du patient sur les rapports de radiographie ou d'imagerie par résonance magnétique (*IRM*) ou dans des fichiers de sons audio médicaux associés au patient. Cette application est importante car il est fortement conseillé d'inscrire le nom du patient sur les rapports, et réduire ainsi les erreurs de classement de ces rapports, ce qui est très important pendant le traitement [66].

#### **3.3.4.5 Autres applications**

Il existe d'autres applications du tatouage numérique audio qui ont été cités dans la littérature. Parmi ces applications on peut citer : la protection contre la copie (copy protection) qui utilise la marque insérée pour empêcher une personne qui détient le document numérique de le dupliquer illégalement et peut aussi être utilisée pour le contrôle de diffusion (broadcast monitoring), une autre application de tatouage audio appelée empreintes digitales (fingerprinting), la marque empreinte digitale insérée est utilisée pour retrouver les utilisateurs autorisés qui violent l'accord de licence de distribution d'un contenu numérique protégé par le droit d'auteur. Dans ce cas la marque insérée dans le contenu concerne le client comme son numéro d'identification.

#### **3.3.5 Selon le processus de détection**

Selon la combinaison des entrées et des sorties du processus de détection, le tatouage est classé comme technique aveugle, semi aveugle ou non aveugle [62].

##### **3.3.5.1 Tatouage non aveugle**

La technique de tatouage non aveugle nécessite le signal hôte d'origine et la marque pour la détection. Il a une robustesse plus forte, mais son application est limitée.

##### **3.3.5.2 Tatouage semi-aveugle**

Cette technique ne nécessite pas le signal hôte pour la détection mais la marque initiale est utilisée pour la comparaison avec la marque extraite.



### 3.3.5.3 Tatouage aveugle

La technique aveugle n'utilise pas le signal hôte pour l'extraction de la marque. Elle a un champ d'application large, mais nécessite une technologie de tatouage plus élevée.

Le tableau 3.2 montre les classes du tatouage numérique selon ces critères de classification.

Tableau 3.2 - Types de tatouage audio basés sur différents critères

N°	Critère	Classification
1	Domaine	<ol style="list-style-type: none"> <li>1. Spatial: LSB, Etalement de spectre (Spread Spectrum)...</li> <li>2. Fréquentiel: DWT, DCT, DFT, SVD</li> <li>3. Cepstral</li> <li>4. Hybride</li> </ol>
2	Technique d'insertion	<ol style="list-style-type: none"> <li>1. Multiplicative</li> <li>2. Substitutive</li> </ol>
3	Perceptibilité	<ol style="list-style-type: none"> <li>1. Robuste: Non affecté par les attaques</li> <li>2. Fragile: Facile à manipuler.</li> <li>3. Semi-Fragile: Résistant pour certains types d'attaques comme la compression sans perte ou l'ajout de bruit</li> </ol>
4	Application	<ol style="list-style-type: none"> <li>1. Protection de droit d'auteur</li> <li>2. Authentification</li> <li>3. Transmission de données</li> <li>4. Autres</li> </ol>
5	Extraction de données tatouées	<ol style="list-style-type: none"> <li>1. Aveugle</li> <li>2. Semi-aveugle</li> <li>3. Non-aveugle</li> </ol>

## 3.4 LES TECHNIQUES FONDAMENTALES DE TATOUAGE AUDIO NUMERIQUE

Le tatouage audio est l'un des domaines les plus difficiles du tatouage, car la détection de la marque dans l'audio est beaucoup plus facile que dans l'image. Comparé au développement du tatouage des images et des vidéos numériques, le tatouage audio numérique représente un défi particulier car (1) les données cachées ne doivent pas être audibles, sinon elles masqueront le signal audio original et seront facilement altérées et supprimées, (2) le système auditif humain (HAS) fonctionne sur une large gamme de fréquences dynamique entre 20 Hz et 20 kHz, ce qui rend difficile l'insertion en dehors de cette gamme, (3) la zone d'insertion des données est limitée.

Le tatouage audio est utilisé pour la sécurité des signaux audio comme la protection des droits d'auteur et l'authentification. Dans un tatouage audio, la marque est insérée dans un signal audio hôte de telle manière que le contenu d'insertion ne soit pas perçu. En même temps, la marque insérée doit être suffisamment robuste contre diverses attaques de tatouage audio. La figure 3.8 illustre le tatouage d'un signal audio initial par un filtre spatial ou un filtre fréquentiel en utilisant dans ce cas une transformation directe et une transformation inverse pour récupérer le signal audio tatoué. Cette section présente les principales techniques de

tatouage audio dans les différents domaines d'insertion, temporel (spatial), fréquentiel, et hybride.

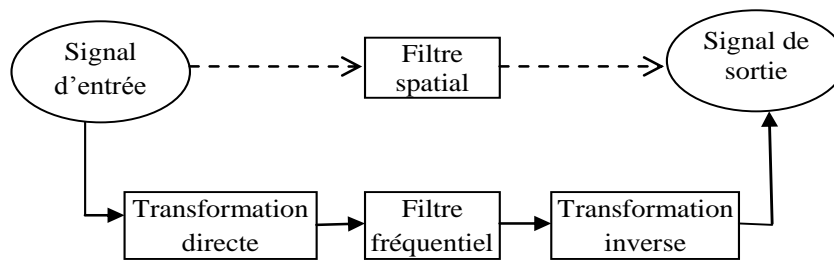


Fig. 3.8 - Domaines d'algorithmes de tatouage numérique

### 3.4.1 Tatouage audio du domaine spatial

Ce type de technique de tatouage modifie les valeurs de l'échantillon du signal audio [67]. En général, ces types de techniques sont simples et n'ont pas besoin du signal audio original pour extraire la marque (tatouage aveugle). Ils offrent également une bonne capacité de charge utile et de transparence. Cependant, ils ont l'inconvénient de ne pas être robustes contre les attaques de tatouage parce que la marque insérée n'est pas distribuée sur le signal audio et les attaques peuvent facilement la déformer. Les algorithmes de tatouage audio spatial sont catégorisés dans l'insertion par substitution ou additive.

Il faut noter que certaines méthodes de tatouage considérées du domaine temporel comme l'étalement du spectre (SS), patchwork et modulation d'indice de quantification (QIM) sont généralement utilisées avec une technique de transformation fréquentielle comme DFT, DCT ou DWT. Certaines techniques de tatouage audio numérique ont été des techniques existantes provenant d'autres domaines de recherche et améliorées pour le tatouage audio, par exemple l'étalement du spectre (SS) de la théorie des communications et les méthodes de patchwork du tatouage d'image. Au début des années 2000, la technique de modulation de l'indice de quantification (QIM), a été introduite pour les formats multimédias généraux et a été utilisée pour le tatouage audio numérique [68].

#### 3.4.1.1 Tatouage Audio par Substitution LSB (Low-Significant-Bit Coding)

C'est l'une des techniques les plus simples pour le tatouage audio [69]–[71]. Dans cette technique, un bit spécifique du signal audio est remplacé par le bit de la marque. L'emplacement du bit spécifique est sélectionné avant l'insertion. A la réception le détecteur connaît la position du bit dans le signal audio tatoué où la marque est insérée. En utilisant cette information, le détecteur extrait le bit tatoué de cette position. Une des approches la plus connue est la substitution de bit le moins significatif (*least significant bit - LSB*) où le bit marqué est inséré dans le dernier bit des échantillons du signal audio pour générer le signal audio tatoué. La technique LSB peut être facilement comprise grâce à l'exemple illustré à la figure 3.9. Supposons que deux valeurs de l'échantillon audio dans le fichier hôte soient 130 (10 000 010) et 150 (10 010 110). Ensuite, en utilisant la technique LSB, si la marque

insérée est 10, alors les valeurs de l'échantillon tatoué seront 131 (10 000 011) et 150 (10 010 110).

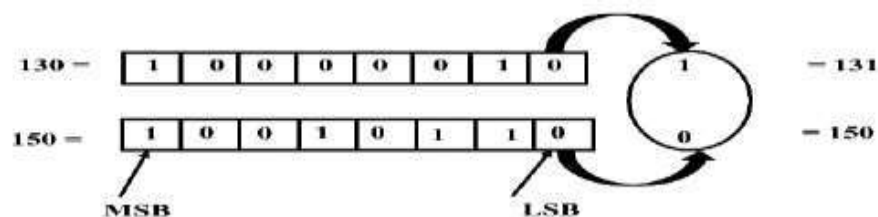


Fig. 3.9 - Exemple de technique de base du bit le moins significatif (LSB)

Côté extraction, l'emplacement spécifique du bit de la marque dans le signal audio tatoué est connu. Par conséquent, tout d'abord, l'échantillon tatoué du signal audio est converti en valeurs de 8 bits. Ensuite, le bit le moins significatif est pris en tant que bit de la marque.

En général, la mise en œuvre de cette technique est très facile et a une capacité de charge utile élevée. Mais cette technique offre moins de robustesse contre diverses attaques telles que la compression avec perte et les processus de traitement du signal. Par conséquent, cette technique n'est pas largement utilisée seule dans le scénario actuel mais ajoutée à d'autres techniques de tatouage audio.

Plusieurs travaux de recherche ont été proposés pour améliorer la méthode LSB. Cvejic et Seppanen dans [72], ont appliqué une stratégie simple de LSB pour insérer un message vocal dans la communication sans fil. Cependant, dans une tentative d'augmenter la capacité de masquage tout en minimisant l'erreur sur le fichier audio, ces auteurs dans [73] ont adopté une méthode de remplacement d'erreur minimum tout en intégrant quatre bits par échantillon. L'erreur d'intégration est ensuite diffusée sur les quatre échantillons suivants. La capacité du canal de données cachées avec cette dernière méthode a atteint 176,3 kbps au signal de 44,1 kHz. Pour améliorer la robustesse contre la distorsion et le bruit de la méthode LSB, Cvejic et Seppanen dans [74], [75] et Ahmed et al. dans [76] ont augmenté la profondeur de la couche d'intégration de la 4ème à la 6ème et 8ème couche LSB sans affecter la transparence perceptuelle du signal audio tatoué. Dans [74], [75] seuls les bits à la sixième position de chaque échantillon de 16 bits du signal hôte d'origine sont remplacés par des bits du message. Pour minimiser l'erreur d'intégration, les autres bits peuvent être retournés afin d'avoir un nouvel échantillon plus proche de celui d'origine. D'autre part, les auteurs dans [76] ont déplacé le LSB de l'intégration au huitième bit et ont évité de cacher dans les périodes silencieuses ou proche des points silencieux dans le signal hôte. Le fait que l'encastrement se produise dans le huitième bit augmentera légèrement la robustesse de cette méthode par rapport aux méthodes LSB conventionnelles. Cependant, la capacité de masquage diminuera puisque certains des échantillons doivent être laissés inchangés pour préserver la qualité sonore perceptuelle du signal audio [77].

### 3.4.1.2 Tatouage audio additif

C'est une autre technique simple de tatouage audio dans le domaine spatial [78], [79]. Dans cette technique, le bit de la marque est inséré dans le signal audio en

utilisant un facteur de mise à l'échelle. Si  $S$  est un signal audio hôte et  $W$  est une marque à insérer, le signal audio  $S^*$  est généré en utilisant la formule suivante :

$$S^* = S + \alpha \times W \quad (3.10)$$

Côté extraction, le signal audio original et le facteur de mise à l'échelle sont nécessaires pour extraire la marque du signal audio tatoué. Le processus d'extraction de la marque est un processus inversé de l'insertion et donné par la formule ci-dessous :

$$W' = (S^* - S)/\alpha \quad (3.11)$$

Où  $W'$  est la marque extraite.

Cette technique est un type de tatouage audio fragile et utilisée pour l'authentification du droit d'auteur du signal audio.

### 3.4.1.3 Masquage de l'écho

La méthode de dissimulation par écho (*Echo Hiding*) insère des données dans le signal audio en introduisant un écho court au signal hôte. La nature de l'écho est une résonance ajoutée à l'audio hôte. Par conséquent, le problème de sensibilité du système auditif humain HAS (*Humain Auditory System*) au bruit additif est évité. Après l'ajout de l'écho, le signal audio tatoué conserve les mêmes caractéristiques statistiques et perceptuelles. C'est une méthode de tatouage additif et aveugle destiné essentiellement aux signaux audio. Un bit est inséré par écho. Pour ajouter plusieurs bits, le signal original est décomposé en petits segments et un bit de la marque est caché dans chaque segment en ajoutant son écho.

Les données sont masquées en manipulant trois paramètres du signal de l'écho : amplitude initiale, décalage (retard) et taux de décroissance de sorte que l'écho ne soit pas audible [78]. Pour un retard allant jusqu'à 1 ms entre le signal original et l'écho, l'effet est indissociable. En plus, l'amplitude et le taux de décroissance pourraient être fixés à des valeurs sous le seuil audible de l'oreille humaine. Dans ce dernier cas, les données pourraient être cachées sans être perceptibles. Cependant, l'inconvénient est un faible taux de détection qui limite l'application. En raison de faible taux d'insertion et d'une sécurité faible, aucun système de tatouage audio basé sur la dissimulation d'écho n'a été présenté dans les recherches récentes. De plus, peu de techniques ont été proposées pour le tatouage audio. Afin d'améliorer la robustesse du système de tatouage contre le traitement commun des signaux, une technique intéressante de propagation de l'écho-temps a été proposée par Erfan et Siahpoush dans [79]. Par rapport au système de masquage d'écho classique, la méthode proposée détecte le bit tatoué basé sur la quantité de corrélation au récepteur et non sur le retard et a un meilleur taux de détection [77].

### 3.4.1.4 Codage de phase

Une donnée cachée est représentée par une phase particulière ou un changement de phase dans le spectre de phase. Le codage de phase (*phase coding*) est l'un des schémas de codage les plus efficaces en termes de rapport signal/bruit, car les expériences indiquent que les auditeurs peuvent ne pas entendre de différence causée par un décalage de phase régulier, même si les modèles de signaux peuvent changer radicalement [63]. L'idée de base de cette technique est de diviser le signal audio,

ensuite substituer la phase de la première trame avec une phase de référence pour cacher les données de la marque. En terme d'imperceptibilité cette méthode est très efficace, mais son inconvénient est la capacité faible d'insertion car seul la première trame est employée pour insérer la marque.

Pour une plus grande capacité d'insertion, et en utilisant des filtres numériques passe-partout (APFs), Ansari et al. dans [80] intègrent des données dans des sous-bandes sélectionnées en utilisant des modèles distincts de APF. Pour une capacité de masquage de 1.2 Kbps, le schéma proposé est robuste contre l'addition de bruit, la compression avec une erreur inférieure à 2%, le découpage aléatoire, la re-quantification et le re-échantillonnage. Pour augmenter encore la robustesse du schéma de masquage de phase, un ensemble de  $n$  APFs d'ordre  $n$  a été utilisé dans [82]. La valeur de  $n$  est un nombre entier positif et les emplacements des pôles peuvent être choisis de diverses façons. Les données sont intégrées dans certains paramètres du APF et récupérées à l'aide du spectre de puissance pour estimer l'emplacement des pôles du APF. Cette méthode résiste aux attaques de manipulation de données standard et réalise une capacité d'intégration de 243 bps tout en fournissant des performances perceptives comparables et une meilleure robustesse [77].

### 3.4.1.5 Étalement de spectre

La technique de base de l'étalement du spectre (*SS-spread spectrum*) est conçue pour cacher un flux d'informations en répartissant les données sur la plus grande partie possible du spectre de fréquences. Il s'avère que certaines techniques d'étalement du spectre s'adaptent bien au masquage de données dans les signaux audio.

Il ya deux approches de base de l'étalement de spectre : *time hopping* (dite aussi *direct sequence*) et *frequency hopping*. Dans l'approche *direct sequence* qui est la plus utilisée, la marque originale  $W_0$  est modulée par une séquence pseudo aléatoire  $W_C$  générée par une clé  $C$ , pour trouver la marque  $W$  de large bande : on dit que le spectre de  $W$  a été étalé. Pour garder  $W$  inaudible, un facteur  $\alpha$  peut être utilisé pour contrôler l'amplitude de  $W$ . Le signal tatoué  $S'$  est ainsi produit en ajoutant  $W$  au signal hôte  $S$ .

Dans le processus de détection, la marque extraite  $W'$  est récupérée en corrélant le signal reçu  $S'$  avec la séquence  $W_C$  utilisée dans l'insertion.

L'étalement du spectre est un concept mis au point dans le domaine des communications pour assurer une récupération adéquate d'un signal envoyé sur un canal bruyant en produisant des copies redondantes du signal de données. Fondamentalement, les données sont multipliées par un code de séquence pseudo aléatoire connu de l'expéditeur et du récepteur [83], puis incorporé dans l'audio de couverture. Ainsi, si le bruit corrompt certaines valeurs, il y aura toujours des copies de chaque valeur restante pour récupérer le message incorporé. *Hernandez\_Garay et al.* [84] ont appliqué la technique conventionnelle de diffusion du spectre en séquence directe (*direct sequence spread spectrum -DSSS*) pour cacher des données confidentielles dans les signaux numériques audio MP3 et WAVE. Pour contrôler la distorsion du fichier audio tatoué, les auteurs dans [85] et [86] ont proposé une

méthode d'insertion où les données étalées sont cachées sous un masque de fréquence. *Matsuka* dans [85], a combiné le spectre étendu au déphasage pour augmenter la robustesse des données transmises contre le bruit additif et permet une détection facile des données incorporées. Dans cette méthode, une capacité de dissimulation fiable de 3 bps a été atteinte. Pour un meilleur taux de dissimulation de 20 bps, *Li et Yu* dans [86] ont utilisé la technique d'étalement de spectre dans le domaine des sous-bande. Des coefficients de sous-bande choisis de façon appropriée ont été sélectionnés pour régler le problème de robustesse et résoudre l'incertitude de synchronisation au décodeur.

Cette technique d'étalement de spectre est un type de tatouage aveugle et robuste. Cependant le processus d'insertion peut introduire une distorsion perceptible sur le signal audio. L'ajustement de l'amplitude par le seuil du modèle psycho acoustique est souvent utilisé pour assurer l'inaudibilité de la marque [77].

### 3.4.1.6 Algorithme de Patchwork

Patchwork est un excellent algorithme de tatouage proposé pour les images. *Bender et al.*, en 1996 [87] ont proposé l'idée de base. Cet algorithme incorpore une statistique spéciale dans une image hôte. Les deux principales étapes de l'algorithme sont les suivantes (i) choisir deux patches de manière pseudo-aléatoire et (ii) ajouter la petite valeur constante  $d$  aux valeurs d'échantillon d'un patch  $A$  et soustraire la même valeur  $d$  des valeurs d'échantillon d'un autre patch  $B$  [88].

$$a_i^* = a_i + d, \quad b_i^* = b_i - d \quad (3.12)$$

Où  $a_i$  et  $b_i$  sont des valeurs d'échantillon des ensembles de patchwork  $A$  et  $B$ , respectivement. Ainsi, les valeurs d'échantillon originales doivent être légèrement modifiées. Le processus de détection commence par la soustraction des valeurs d'échantillon entre deux patches. Alors,  $E[\bar{a}^* - \bar{b}^*]$  la valeur attendue des différences des moyennes des échantillons est utilisée pour décider si les échantillons contiennent ou non des informations de marque, où  $a^*$  et  $b^*$  sont les moyennes d'échantillon de l'échantillon individuel  $a_i^*$  et  $b_i^*$ , respectivement. Puisque deux patches sont utilisés au lieu d'un, on peut détecter les filigranes insérés sans les images hôtes originales, ce qui en fait un algorithme de tatouage aveugle. Le patchwork est en soi un très bon algorithme, mais il présente certains inconvénients. Il faut noter que :

$$E[\bar{a}^* - \bar{b}^*] = E[(\bar{a} + d) - (\bar{b} - d)] = E[\bar{a} - \bar{b}] + 2d \quad (3.13)$$

Où  $\bar{a}$  et  $\bar{b}$  sont les moyennes d'échantillon de l'échantillon individuel  $a_i$  et  $b_i$ , respectivement. L'algorithme du patchwork suppose que  $E[\bar{a}^* - \bar{b}^*] = 2d$ , en raison de l'hypothèse préalable selon laquelle l'échantillon aléatoire garantit que les valeurs attendues sont toutes les mêmes, de sorte que  $E[\bar{a}^* - \bar{b}^*] = 0$ . Cependant, la différence réelle des moyennes de l'échantillon,  $\bar{a} - \bar{b}$  n'est pas toujours nulle dans la pratique. Bien que la distribution tatouée soit décalée vers la droite comme le montre la figure. 2.13, la probabilité d'une détection erronée existe, la zone étant inférieure à 0 dans la distribution tatouée. Les performances de l'algorithme de patchwork dépendent de la distance entre deux moyennes des échantillons et  $d$  ce qui affecte l'inaudibilité. En plus, l'algorithme du patchwork a été conçu à l'origine pour

les images mais l'image diffère de l'audio dans beaucoup de points. Ainsi, il n'est pas pratique d'appliquer cet algorithme directement au tatouage audio [88].

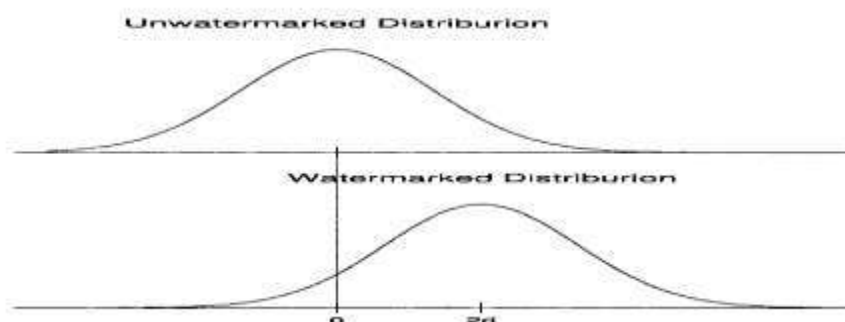


Fig. 3.10 - Comparaison des distributions non tatouées et tatouées de la différence de moyenne

L'algorithme original du patchwork a été appliqué aux données du domaine spatial (ou, du domaine temporel en audio). Cependant, l'intégration dans le domaine temporel est vulnérable même aux attaques et aux modifications de faible ampleur. En fait, Arnold dans [89] a appliqué l'algorithme du patchwork dans le domaine des fréquences, mais pas dans le domaine temporel. Une autre contribution importante de son travail a été l'introduction de la fonction de densité de puissance dans les tests d'hypothèse. En plus, son algorithme diffère de l'algorithme original de patchwork dans plusieurs points :

1. La moyenne et la variance des valeurs de l'échantillon sont calculées afin de détecter les marques. Dans l'algorithme original [87], seule la moyenne était utilisée.
2. Arnold [89] suppose que la distribution des valeurs de l'échantillon est normale. L'algorithme original [87] suppose que la distribution est uniforme.
3. La fonction d'intégration d'Arnold est multiplicative, comme  $a_i^* = a_i(1 + d)$ ,  $b_i^* = b_i(1 - d)$ . L'algorithme original [87] était basé sur la fonction d'insertion additive.
4. Arnold [89] essaie de décider de la valeur  $d$  de manière adaptative. Cependant, l'algorithme original [87] a choisi une valeur fixe de 0,15 dans l'expérimentation.

La fonction d'insertion donne lieu à des changements considérables du facteur d'échelle lorsque  $d$  est grand puisqu'elle est multiplicative. Donc une sérieuse dégradation de la qualité est attendue. La valeur attendue des moyennes de l'échantillon est donnée comme suit :

$$[\bar{a}^* - \bar{b}^*] = E[\bar{a}(1 + d) - \bar{b}(1 - d)] = E[\bar{a} - \bar{b}] + dE[\bar{a} + \bar{b}] \quad (3.14)$$

Si  $E[\bar{a} + \bar{b}] = 0$ , alors  $E[\bar{a}^* - \bar{b}^*] = 0$  et, par conséquent, il ne peut pas contribuer à élargir l'écart entre les deux pics indiqués à la figure 3.10. Cela signifie que les pics sont trop rapprochés pour détecter les filigranes. Afin d'élargir l'écart, la valeur doit être grande. Malheureusement, cela rend la qualité audio plus mauvaise. Ainsi, il est nécessaire que  $E[\bar{a} + \bar{b}] \neq 0$  pour une application réussie. Cependant, cette exigence peut parfois ne pas être respectée, par exemple dans le domaine DCT. Une autre contribution d'Arnold [89] est l'idée de décider que la valeur de  $d$  soit adaptative. Il s'agit d'un concept très important car il peut permettre à l'algorithme de

tatouage de cacher des informations en fonction des caractéristiques de l'audio. Ainsi, l'inaudibilité et la robustesse peuvent être considérablement améliorées [88].

### 3.4.1.7 La modulation de l'indice de quantification (QIM- Quantization index modulation)

La technique de modulation de l'indice de quantification (QIM) a été proposée par Chen et Wornell comme méthode de tatouage numérique valide [90]. Elle présente un très bon compromis robustesse-transparence et cache des informations en quantifiant des échantillons. Dans la méthode QIM, chaque échantillon est quantifié avec un pas de quantification prédéterminé. Ensuite, une légère modification est apportée à chaque échantillon quantifié en fonction des valeurs des bits de la marque. Dans [91], l'auteur a introduit une implémentation simple de QIM comme suit : supposons que l'échantillon hôte d'entrée est  $X$ , le pas de quantification est  $\Delta$ , et le bit de la marque à intégrer est  $w \in \{0,1\}$ , alors l'échantillon tatoué  $\hat{X}$  est obtenu comme suit

$$\hat{X} = Q(X, \Delta) + \frac{(2w-1)\Delta}{4} \quad (3.15)$$

où  $Q(X, \Delta)$  est la fonction de quantification standard donnée par

$$Q(X, \Delta) = \left\lfloor \frac{X}{\Delta} \right\rfloor \Delta \quad (3.16)$$

Avec  $[\cdot]$  désignant l'arrondi d'une valeur au nombre entier le plus proche.

La figure 3.11 illustre l'insertion de la marque de cette méthode. L'échantillon  $X$  est d'abord quantifié à la valeur  $Q(X, \Delta)$  (cercle noir). Si le bit de la marque à intégrer est 1, alors  $\Delta/4$  est ajouté à la valeur de l'échantillon quantifié, ce qui fait monter l'échantillon jusqu'au cercle pendant. Sinon,  $\Delta/4$  est soustrait de la valeur de l'échantillon quantifié, ce qui fait descendre l'échantillon vers la croix.

Du côté du décodeur, la valeur absolue de la différence entre l'échantillon reçu et sa valeur quantifiée est calculée. Si elle est comprise entre 0 et  $\Delta/4$ , alors le bit de la marque extraite est "1". Si la différence est comprise entre  $-3\Delta/4$  et 0, alors le bit de la marque extraite est "0". Sinon, le signal reçu n'est pas tatoué. Ceci peut être illustré comme suit

$$w = \begin{cases} 1, & 0 \leq |\hat{X} - Q(\hat{X}, \Delta)| < \frac{\Delta}{4} \\ 0, & \frac{\Delta}{4} \leq |\hat{X} - Q(\hat{X}, \Delta)| \leq \Delta \end{cases} \quad (3.17)$$

De manière équivalente, par le décodeur à distance minimale, comme suit

$$w = \underset{\omega \in \{0,1\}}{\operatorname{argmin}} \left| Q(\hat{X}, \Delta) + \frac{(2\omega - 1)\Delta}{4} - \hat{X} \right| \quad (3.18)$$

QIM est très simple à mettre en œuvre, optimal en termes de capacité de canal et robuste aux attaques par ajout de bruit. Tant que le bruit introduit dans le canal de transmission est inférieur à  $\Delta/4$ , le détecteur peut toujours extraire correctement le bit de la marque. Cependant, si le bruit dépasse  $\Delta/4$ , le bit de la marque peut ne pas être détecté avec précision. Il s'agit d'un compromis entre la robustesse de la marque et la transparence. Un pas de quantification plus grand  $\Delta$  conduit à une plus grande





audio et l'HAS. Avec le schéma d'insertion adaptatif au contenu, le processus d'insertion varie en fonction du contenu du signal audio. En gros, cette technique consiste à segmenter un signal audio en trames dans le domaine temporel, à classer les trames comme appartenant à l'une de classes connues, puis à coder chaque trame avec un schéma d'insertion approprié. Pour améliorer la robustesse et la résistance à l'effacement de la marque intégrée, une technique de saut à plusieurs bits est utilisée. Au lieu d'insérer un bit dans une trame audio, plusieurs bits avec des délais différents peuvent être insérés dans chaque sous-trame audio [63].

#### **3.4.1.10 Technique de tatouage audio basée sur les périodes de silence**

Cette approche de tatouage audio présente certaines similitudes avec le tatouage textuel à espace ouvert. Les contreparties audio des espaces de texte sont les périodes de silence. Dans un signal audio, une période de silence offre les avantages suivants : (1) une période de silence fait partie intégrante de tout signal audio. Elle ne peut être omise car elle représente une partie intégrante du signal studio. (2) Elle se produit aléatoirement dans le signal audio, qui est généré par le processus musical. (3) Une période de silence représente un intervalle de temps réel qui ne sera pas diminué lors de la compression. La technique présentée dans ce schéma est une technique de tatouage basée sur les périodes de silence sourdine, qui offre les caractéristiques suivantes : (1) elle prolonge les périodes de silence dans un signal audio sans aucune différence de perception pour le système auditif humain moyen. (2) L'extension des périodes de silence a la même amplitude, de sorte qu'elle se fonde avec l'original et n'attire pas l'attention. (3) Il n'est pas nécessaire d'avoir le signal original pour extraire la marque [63].

#### **3.4.2 Tatouage audio du domaine de transformation**

Le tatouage audio de domaine de transformation offre une meilleure robustesse par rapport au tatouage audio de domaine spatial. Dans cette approche, le signal audio se convertit en ses coefficients de transformation, et la marque est insérée à ces coefficients pour obtenir des coefficients modifiés. Ensuite, la transformation inverse est appliquée à ces coefficients modifiés pour obtenir le signal audio tatoué. Les fameuses transformations de signaux telles que les transformées de Fourier, la transformée en cosinus discrète (DCT), la transformée en ondelettes discrète (DWT) et la décomposition en valeurs singulières (SVD) sont principalement utilisées [93]–[95]. La figure 3.12 décrit la procédure utilisée par cette technique.

Les algorithmes de tatouage audio du domaine de transformation sont divisés en deux types, selon la méthode d'insertion de la marque : tatouage par substitution de coefficients, et tatouage aditif ou multiplicatif. La particularité de ces méthodes est de fournir une plus grande imperceptibilité et une plus grande robustesse aux manipulations du média numérique et aux attaques courantes de traitement du signal, mais le coût de calcul est plus élevé que celui des méthodes de tatouage dans le domaine spatial. Ces méthodes ne peuvent pas établir automatiquement un équilibre entre l'imperceptibilité, la robustesse et la capacité, un compromis entre ces trois propriétés doit être pris en compte durant la conception et l'implémentation du système de tatouage.

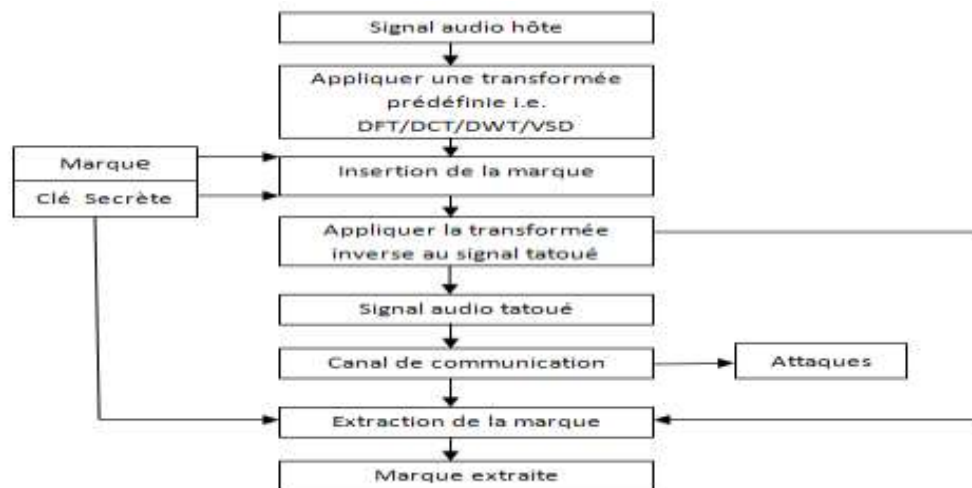


Fig. 3.12 - Insertion et extraction de la marque dans le domaine de transformation

### 3.4.2.1 Les transformées de Fourier

Dans cette section, les concepts de transformée de Fourier discrète, de transformée de Fourier fractionnaire, de transformée de Fourier discrète quaternaire sont brièvement décrits.

#### 1) Transformée de Fourier discrète -DFT

La transformée de Fourier consiste à décomposer un signal ou une image en une somme de signaux élémentaires, qui ont la propriété d'être faciles à mettre en œuvre et à observer [96]. Ces signaux élémentaires sont périodiques et complexes, afin de permettre une étude d'amplitude et de phase des systèmes [97]. La transformée de Fourier discrète d'un signal  $f(m, n)$  de taille  $M \times N$  est définie par :

$$F(u, v) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) e^{-j2\pi \left( \frac{u \cdot m}{M} + \frac{v \cdot n}{N} \right)} \quad (3.19)$$

Où  $u$  et  $v$  désignent les fréquences spatiales pour une position  $m$  et  $n$  à partir de la transformée de Fourier.

Le résultat obtenu après la transformée de Fourier discrète est complexe [98]. En général, pour représenter cette forme complexe, on calcule le module  $M$  et la phase  $P$  donnés par les équations suivantes :

$$M(u, v) = |F(u, v)| = \sqrt{\text{Re}(u, v)^2 + \text{Im}(u, v)^2} \quad (3.20)$$

$$P(u, v) = \arg(F(u, v)) = \tan^{-1} \left( \frac{\text{Im}(u, v)}{\text{Re}(u, v)} \right) \quad (3.21)$$

Où  $\text{Im}$  et  $\text{Re}$  représentent respectivement la partie imaginaire et la partie réelle de la transformée de Fourier du signal  $f$ .

La figure 3.13 montre la distribution des fréquences dans le module de la transformée de Fourier. En raison des propriétés de symétrie hermitienne de la

transformée de Fourier, les modules des premier et deuxième quadrants sont égaux aux coefficients des quatrième et troisième quadrants respectivement.

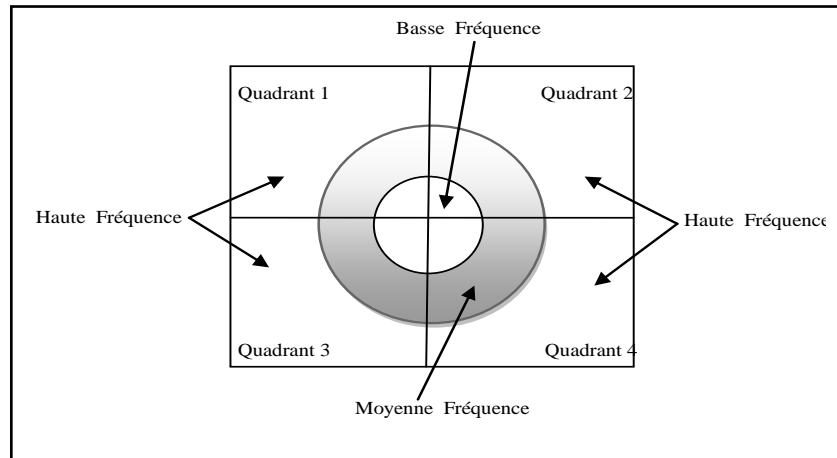


Fig. 3.13 - Distribution de fréquences dans le module de la transformée de Fourier

## 2) Transformée de Fourier fractionnaire -FRFT

La transformée de Fourier fractionnaire (FRFT) est largement utilisée dans le traitement du signal, elle est considérée comme une forme généralisée de la transformée de Fourier [99], elle permet d'obtenir les composantes mixtes temps et fréquence du signal [100]. La FRFT d'une fonction  $f(x)$  avec un ordre fractionnaire  $\alpha$  peut être calculée comme suit :

$$F^\alpha\{f(x)\} = \int_{-\infty}^{+\infty} f(x)K_\alpha(x,u)dx \quad (3.22)$$

Où  $K_\alpha(x,u)$  est le noyau de transformation de la FrFT et est donné par

$$F^\alpha\{f(x)\} = \int_{-\infty}^{+\infty} f(x)K_\alpha(x,u)dx \quad (3.23)$$

$$K_\alpha(x,u) = \sqrt{1 - i \cot \alpha} \exp \left[ i\pi \left( \frac{x^2+u^2}{\tan \alpha} - \frac{2xu}{\sin \alpha} \right) \right] \quad (3.24)$$

La Transformée de Fourier Fractionnaire combine le domaine spatial et le domaine fréquentiel selon un opérateur  $\alpha$ , si  $\alpha$  est proche de 0 la transformée est considérée comme dominante dans le domaine temporel, pour un  $\alpha$  proche de 1 la transformée est considérée comme dominante dans le domaine fréquentiel [101]. L'évolution du domaine spatial vers le domaine fréquentiel (avec  $\alpha$  allant de 0 à 1) permet d'avoir les doubles caractéristiques du domaine temporel et de fréquence.

## 3) Transformée de Fourier discrète en quaternion-QDFT

La QDFT est une généralisation de la DFT complexe bidimensionnelle. Dans une QDFT comme dans une transformée de Fourier discrète bidimensionnelle, le positionnement des points dans le domaine spectral en fonction de la fréquence spatiale horizontale et verticale est le même [102]. Dans le format habituel d'un tableau bidimensionnel de points spectraux, le point dans le coin supérieur gauche représente la fréquence zéro et les fréquences spatiales plus élevées sont représentées

plus près de l'axe horizontal ou vertical du tableau. La QDFT est obtenue en calculant la formule suivante :

$$F(u, v) = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-\mu 2\pi \left(\frac{mu}{M} + \frac{nv}{N}\right)} f(m, n) \quad (3.25)$$

Où  $f(m, n)$  est un signal de taille  $M \times N$  et  $\mu$  est un quaternion pur unitaire.

Au contraire à la DFT bidimensionnelle, qui fusionne les composantes en une seule composante réelle ou imaginaire [103], la QDFT sépare les quatre combinaisons possibles des composants cosinus et sinus horizontale et verticale en quatre composantes du point spectral quaternaire.

#### 4) *Caractéristiques des transformées de Fourier*

Le domaine DFT a été exploré par les chercheurs parce qu'il offre une robustesse contre les attaques géométriques comme la rotation, la mise à l'échelle, le recadrage, la translation, etc. Les principales caractéristiques de la Transformée de Fourier Discrète sont [59], [60]:

- La DFT d'un signal hôte est généralement à valeur complexe, ce qui entraîne une représentation de la phase et de l'amplitude du signal.
- La DFT est invariante à la translation. Les décalages spatiaux du signal affecte la représentation de la phase mais pas la représentation de l'amplitude, ou des décalages circulaires dans le domaine spatial n'affectent pas la magnitude de la transformée de Fourier.
- La DFT est également résistante au recadrage, car l'effet du recadrage entraîne un flou du spectre. Si les marques sont incorporées dans la magnitude, qui correspond à des coordonnées normalisées, il n'y a pas besoin de synchronisation.
- Les composantes les plus fortes de la DFT sont les composantes centrales qui contiennent les basses fréquences.
- La mise à l'échelle du signal entraîne l'amplification du signal extrait et peut être détecté par le coefficient de corrélation. La translation du signal n'a aucun effet sur le signal extrait.
- La rotation du signal entraîne des décalages cycliques du signal extrait et peut être détectée par une recherche exhaustive.
- Un changement d'échelle dans le domaine spatial entraîne un changement d'échelle inverse dans le domaine des fréquences. Une rotation dans le domaine spatial entraîne la même rotation dans le domaine des fréquences.

#### 5) *Critères de sélection des coefficients*

- La modification des coefficients de basse fréquence peut provoquer des distorsions visibles dans le domaine spatial. Par conséquent les coefficients de basse fréquence doivent être évités.
- Les coefficients de haute fréquence ne sont pas appropriés car ils sont supprimés pendant la compression MP3 (compression avec perte).
- Le meilleur emplacement pour insérer la marque est la bande de fréquence moyenne [59].

#### 6) *Avantage de la DFT par rapport aux autres transformées*

- La DFT est invariante en termes de rotation, de mise à l'échelle et de translation (RST). Elle peut donc être utilisée pour récupérer des distorsions géométriques, alors que dans le domaine

spatial, les autres transformées ne sont pas invariantes par rapport à la rotation, l'échelle et la translation et il est donc difficile de surmonter les distorsions géométriques [60].

### 3.4.2.2 La Transformée en Cosinus Discrète – DCT

La transformée en cosinus discrète convertit le signal du domaine temporel en signal du domaine fréquentiel. La DCT à deux dimensions donne les coefficients de fréquence sous la forme d'une matrice. Le coin supérieur gauche de la matrice représente les coefficients de fréquence les plus bas, tandis que le coin inférieur droit représente les coefficients de fréquence les plus élevés. En utilisant la DCT, le média hôte est divisé en pseudo bandes de fréquences, et généralement la marque est insérée dans la sous-bandes de fréquence moyenne. Inversement, si la marque est insérée dans la sous bande de hautes fréquences, elle est plus facile à cacher, mais le schéma de tatouage est moins résistant aux attaques [104]–[106].

Cette importante transformation a été initiée par (Ahmed et al., 1974) [107] et a été utilisée et étudiée largement car elle est très importante pour la compression de donnée. L'un des premiers algorithmes présentés par (Cox et al., 1997) [108] a utilisé l'approche DCT globale pour intégrer une marque robuste dans la partie perceptivement significative du système visuel humain (HVS). Comme d'autres transformations, la transformée en cosinus discrète (DCT) tente de décorréler les données du signal. Après décorrélation, chaque coefficient de transformation peut être codé de manière indépendante sans perte d'efficacité de compression.

La transformée en cosinus discrète (DCT) est une transformation spectrale, qui a les propriétés de la transformée de Fourier discrète [109], mais qui n'utilise que des fonctions cosinus de différents nombres d'ondes comme des fonctions de base. Elle fonctionne sur des signaux à valeurs réelles ou sur des coefficients spectraux. Les définitions de la DCT et de la DCT inverse sont données par les expressions suivantes [110], [111] :

$$y(k) = w(k) \sum_{n=0}^{N-1} x(n) \cos \pi(2n-1)/2n \quad \text{pour } k = 0, 1, 2, \dots, N-1. \quad (3.26)$$

De même, la transformation inverse est définie comme

$$x(k) = \sum_{n=0}^{N-1} w(n) y(n) \cos \pi(2n-1)/2n \quad \text{pour } k = 0, 1, 2, \dots, N-1. \quad (3.27)$$

Avec :

$$w(k) = \begin{cases} \frac{1}{\sqrt{n}} & \text{pour } k = 0 \\ \sqrt{2/n} & \text{pour } 1 \leq k \leq n \end{cases} \quad (3.28)$$

Ainsi, le premier coefficient de transformation correspond à la composante continue du signal (coefficient DC). Tous les autres coefficients de transformation sont appelés coefficients AC. Dans un signal, la majorité des informations sont concentrées dans les basses fréquences. La transformation du signal par la DCT va

donc faire apparaître une occupation spectrale réduite, où une zone de taille relativement petite, code une grande partie de l'information.

De nombreuses études ont déjà été réalisées sur les méthodes de tatouage numériques dans le domaine DCT. La technique de tatouage DCT fonctionne selon les étapes suivantes :

1. Segmenter le signal en trames non chevauchantes de  $8 \times 8$
2. Appliquer certains critères de sélection de trames (selon les critères de visibilité du système visuel/auditif humain)
3. Appliquer une DCT directe à chacune de ces trames
4. Appliquer les critères de sélection du coefficient (p. ex., le plus élevé)
5. Intégrer la marque en modifiant les coefficients sélectionnés
6. Appliquer la transformation DCT inverse sur chaque trame

Cette technique peut être décrite par la figure 3.14 [111]. Il faut noter que l'extraction de la marque peut se faire de manière inverse.

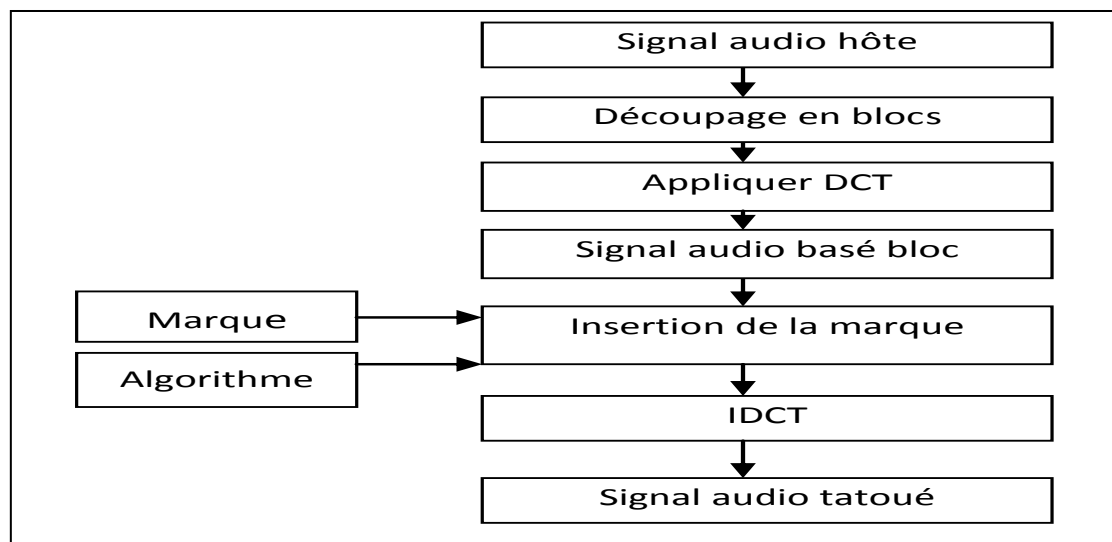


Fig. 3.14 - Insertion de la marque dans un domaine DCT basé trame

La plupart des algorithmes DCT sont classés sur la base des étapes 3 et 4, c'est-à-dire que la principale différence entre la plupart des algorithmes est qu'ils diffèrent soit dans les critères de sélection des trames ou les critères de sélection des coefficients. En fonction de la stratégie de modélisation perceptuelle incorporée par les algorithmes de tatouage, nous trouvons dans la littérature trois classes de ces algorithmes [59] :

1. Des algorithmes basés DCT qui n'intègrent aucune stratégie de modélisation perceptuelle lors de l'intégration d'une marque.
2. Des algorithmes basés DCT qui sélectionnent les coefficients de transformation qui ont une grande capacité de perception ou les coefficients qui sont les moins modifiés par les attaques courantes de traitement du signal comme le filtrage passe-bas et l'ajout de bruit. La composante DC et les coefficients AC ayant une amplitude élevée satisfont ce critère et peuvent être sélectionnés. La composante haute fréquence sont affectées par les

opérations courantes de traitement de l'audio comme la compression, elles ne sont donc pas un bon choix pour le tatouage.

3. Des algorithmes qui intègrent les propriétés du système visuel/auditif humain pour une modélisation perceptive ou inaudible. Seuls les coefficients qui satisfont ce critère sont sélectionnés et modifiés.

### ***1) Propriétés de la DCT pour le traitement de signal***

1. **Décorrélation** : Comme indiqué précédemment, le principal avantage de la transformation du signal est l'élimination de la redondance entre les points voisins. Cela conduit à des coefficients de transformation non corrélés qui peuvent être codés de façon indépendante.
2. **Compactage de l'énergie** : L'efficacité d'un système de transformation peut être directement mesurée par sa capacité à emballer des données dans peu de coefficients que possible. Cela permet au quantificateur de supprimer les coefficients d'amplitude relativement faibles sans introduire une distorsion audible dans le signal reconstruit.
3. **Utilisation de la DCT en tatouage numérique** : La DCT, comme la transformée de Fourier, représente les données en termes d'espace de fréquence plutôt que d'espace d'amplitude. Cette représentation est utile car elle correspond à la façon dont les humains perçoivent la lumière, de sorte que la partie qui n'est pas perçue peut être identifiée et éliminée. Les techniques de tatouage basées sur la DCT sont robustes par rapport aux techniques du domaine spatial. Ces algorithmes résistent aux opérations simples de traitement de signal comme le filtrage passe-bas, l'ajustement de la luminosité et du contraste, le flou, etc. Cependant, ils sont difficiles à mettre en œuvre et sont plus coûteux en termes de calcul. En même temps, ils sont faibles contre les attaques géométriques telles que la rotation, la mise à l'échelle, le recadrage, etc. L'insertion de la marque dans la partie perceptiblement significative du signal a ses propres avantages car la plupart des schémas de compression suppriment la partie perceptiblement insignifiante du signal [60]. Dans le domaine spatial, la partie insignifiante du signal représente le LSB cependant, dans le domaine de la fréquence, elle représente la composante haute fréquence [59].

### **3.4.2.3 La Transformée en ondelette Discrète – DWT**

La transformée en ondelettes est une technique moderne fréquemment utilisée dans le traitement des images numériques, la compression, le tatouage, etc. Au cours des dernières années, la transformée en ondelettes a été largement étudiée dans le traitement du signal en général et la compression d'images en particulier. Dans certaines applications, les systèmes de tatouage basés sur les ondelettes surpassent les approches basées sur la DCT.

Dans l'analyse par ondelettes, la transformée en ondelettes discrètes (DWT) décompose un signal en un ensemble de fonctions de base d'ondelettes mutuellement orthogonales. Ces fonctions diffèrent des fonctions de base sinusoïdales en ce qu'elles sont localisées dans l'espace, c'est-à-dire non nulles sur une partie seulement de la longueur totale du signal. En outre, les fonctions d'ondelettes sont des versions dilatées, translatées et mises à l'échelle d'une fonction commune  $\phi$ , connue comme l'ondelette mère. Comme c'est le cas dans l'analyse de Fourier, la DWT est inversible, de sorte que le signal original peut être complètement récupéré à partir de sa représentation DWT.



Contrairement à la DFT, la DWT en fait, ne fait pas référence à une seule transformation, mais plutôt à un ensemble de transformations, chacune avec un ensemble différent de fonctions de base d'ondelettes. Les ondelettes de Haar et l'ensemble d'ondelettes de Daubechies sont parmi les plus courantes [59].

En mathématiques, une transformée en ondelettes discrètes (DWT) est une transformée qui décompose un signal en ondelettes, plutôt qu'en fréquences. Dans une DWT, les ondelettes sont échantillonnées de manière discrète. La résolution temporelle est l'un des avantages de la DWT par rapport aux transformées de Fourier (c'est-à-dire la DCT et la DFT). Cela fait de la DWT un domaine de recherche plus attractif, en montrant plusieurs aspects de l'information, tels que la localisation dans le temps et la fréquence. Un ensemble d'ondelettes, qui sont des fonctions mathématiques, est utilisé pour décomposer le signal [111].

La transformée en ondelettes est utile dans le traitement des signaux numériques, la compression d'images, et pour éliminer le bruit du signal. L'idée principale d'une transformée en ondelettes est l'utilisation d'un ensemble de fonctions de base (appelées ondelettes) qui offrent une localisation dans le domaine des fréquences. L'utilisation d'une transformée en ondelettes permet d'obtenir une haute résolution fréquentielle à basse fréquence et une haute résolution temporelle à haute fréquence [111].

La transformée discrète en ondelettes (DWT) d'un signal  $x(n)$  est donnée par les deux transformées DWT suivantes [111]:

$$W_\phi(j_0, k) = \frac{1}{\sqrt{M}} \sum_n x(n) \phi_{j_0, k}(n) \quad (3.29)$$

$$W_\psi(j, k) = \frac{1}{\sqrt{M}} \sum_n x(n) \psi_{j, k}(n), \text{ pour } j \geq j_0 \quad (3.30)$$

Où  $W_\phi(j_0, k)$  sont les coefficients d'approximation (AC), et  $W_\psi(j, k)$  sont les coefficients de détail (DC), et la fonction inverse (IDWT) est donnée par [111]:

$$x(n) = \frac{1}{\sqrt{M}} \sum_n W_\phi(j_0, k) \phi_{j_0, k}(n) + \frac{1}{\sqrt{M}} \sum_n W_\psi(j, k) \psi_{j, k}(n) \quad (3.31)$$

$$\begin{aligned} \text{avec } n &= 0, 1, 2, \dots, M-1, & j &= 0, 1, 2, \dots, J-1, \\ k &= 0, 1, 2, \dots, 2^j - 1 \end{aligned} \quad (3.32)$$

Où  $M$  est le nombre d'échantillons à transférer  $= 2^J$ ,  $J$  est le nombre de niveaux de transformation,  $\phi_{j_0, k}(n)$  et  $\psi_{j, k}(n)$  sont deux fonctions de base,  $\phi(n)$  représente la fonction de mise à l'échelle et  $\psi(n)$  représente la fonction d'ondelette.

La transformée en ondelettes décompose le signal en trois directions spatiales, à savoir horizontale, verticale et diagonale. Les ondelettes reflètent donc plus précisément les propriétés du système visuel/auditif humain. La décomposition à un seul niveau donne quatre représentations de fréquences du signal. Ces quatre

représentations sont appelées les sous-bandes LL, LH, HL et HH, comme le montre la figure 3.15.

<b>LL<sub>1</sub></b>	<b>HL<sub>1</sub></b>
<b>LH<sub>1</sub></b>	<b>HH<sub>1</sub></b>

Fig. 3.15 - Décomposition DWT à un seul niveau

L'amplitude des coefficients DWT est plus grande dans les bandes les plus basses (LL) à chaque niveau de décomposition, et plus petite pour les autres bandes (HH, LH et HL). Les sous-bandes LH, HH et HL, sont utilisées pour intégrer la marque. Les sous-bandes sont constituées d'une large gamme du spectre de fréquences du signal hôte. Par conséquent, la robustesse du système de tatouage est augmentée.

La sous-bande LL de basse fréquence peut être décomposée à un autre niveau de décomposition et ce processus peut être appliqué jusqu'à ce que l'obtention d'un niveau de décomposition souhaité déterminé par l'application. La figure 3.16 montre une transformée DWT à trois niveaux [63], [112].

<b>LL<sub>3</sub></b>	<b>HL<sub>3</sub></b>	<b>HL<sub>2</sub></b>	<b>HL<sub>1</sub></b>
<b>LH<sub>3</sub></b>	<b>HH<sub>3</sub></b>		
<b>LH<sub>2</sub></b>		<b>HH<sub>2</sub></b>	<b>HH<sub>1</sub></b>
<b>LH<sub>1</sub></b>			

Fig. 3.16 - Transformée en ondelettes à trois niveaux

Après avoir appliqué la DWT, le système incorpore la marque dans le signal hôte à l'aide d'un algorithme, puis applique la DWT inverse (IDWT) pour obtenir le média hôte tatoué. Le processus d'extraction de la marque prend le signal tatoué en entrée et applique la DWT au même niveau. L'ensemble du processus est décrit dans la figure 3.17 [111].

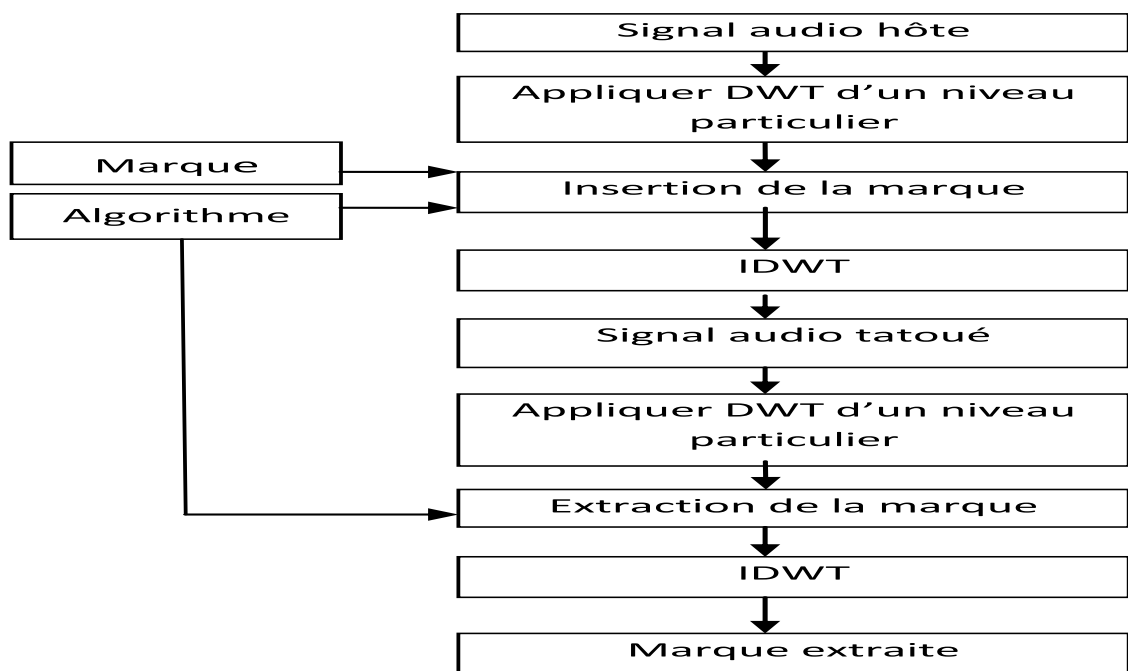


Fig. 3.17 - Insertion et extraction de la marque dans un domaine de transformation DWT

La transformée en ondelettes discrètes (DWT) est actuellement utilisée dans un grand nombre d'applications de traitement du signal, comme la compression audio et vidéo, l'élimination du bruit dans l'audio et la simulation de la distribution des antennes sans fil. Les ondelettes ont leur énergie concentrée dans le temps et sont bien adaptées à l'analyse des signaux transitoires et variables dans le temps. Comme la plupart des signaux rencontrés dans la vie réelle sont de nature variable dans le temps, la transformée en ondelettes convient très bien à de nombreuses applications [63].

### 1) Propriétés de la DWT par rapport au traitement de signal

- La transformée en ondelettes décompose le fichier média hôte dans trois directions spatiales, à savoir l'horizontale, la verticale et la diagonale. Par conséquent, les ondelettes reflètent plus précisément les propriétés anisotropes du système visuel/auditif humain [59].
- La transformation en ondelettes est efficace sur le plan informatique et peut être mise en œuvre en utilisant une simple convolution de filtres.
- L'amplitude des coefficients DWT est plus grande dans les bandes les plus basses (LL) à chaque niveau de la transformation en ondelettes et est plus petite pour les autres bandes (HH, LH, HL) [59].
- Plus l'amplitude du coefficient d'ondelette est grande, plus il est significatif.
- La détection de la marque à des résolutions inférieures est efficace en termes de calcul car à chaque niveau de résolution successif, peu de bandes de fréquences sont concernées.
- Pour le média image, les sous-bandes à haute résolution permettent de localiser facilement les bords et les textures d'une image [59].

## 2) *Avantages de la DWT par rapport à la DCT*

- La transformée en ondelettes comprend le HVS/HAS plus étroitement que la DCT.
- Le média numérique codé par ondelettes est une description multi résolution de l'image. Par conséquent, le média peut être affiché à différents niveaux de résolution et peut être traité séquentiellement de la basse à la haute résolution.
- Les artefacts visuels introduits par les images codées par ondelettes sont moins évidents que ceux introduits par la DCT, car la transformée en ondelettes ne décompose pas l'image en blocs. À des taux de compression élevés, des artefacts de blocage sont perceptibles dans la DCT, mais dans les images codées par ondelettes, ils sont beaucoup plus clairs.
- La DFT et la DCT sont des transformées de trame complète, et donc tout changement dans les coefficients de transformation affecte l'ensemble de l'image, sauf si la DCT est mise en œuvre à l'aide d'une approche par blocs. Cependant, la DWT a une localité de fréquence spatiale, ce qui signifie que si un signal est incorporé, il affectera l'image localement [113]. Par conséquent, une transformée en ondelettes fournit une description à la fois fréquentielle et spatiale d'un signal.

## 3) *Inconvénients de la DWT par rapport à la DCT*

- La complexité de calcul de la DWT est supérieure à celle de la DCT [114]. Comme l'a souligné Feig (1990), dans [115] il suffit de 54 multiplications pour calculer la DCT pour un bloc de 8x8, contrairement au calcul des ondelettes qui dépend de la longueur du filtre utilisé, qui est d'au moins 1 multiplication par coefficient.

### 3.4.2.4 Décomposition en valeur singulière (Singular Value Decomposition-SVD)

En mathématiques, une décomposition en valeurs singulières (SVD) est le produit d'une matrice réelle ou complexe. Cette méthode est la généralisation de la décomposition des valeurs propres d'une matrice symétrique avec des valeurs propres non négatives à toute matrice  $m \times n$  par une extension de la décomposition polaire. La transformation SVD a été largement utilisée en statistique et en traitement du signal numérique. La SVD d'une matrice  $M$  peut être définie par l'équation suivante [111] :

$$M = USV^* \tag{3.33}$$

Où  $M$  est une matrice  $m \times n$  qui provient du champ  $K$  (soit des nombres réels, soit des nombres complexes),  $U$  est une matrice unitaire  $m \times m$  sur  $K$ , ou une matrice orthogonale (si  $K = R$ ),  $S$  est une matrice diagonale  $m \times n$  avec des nombres réels non négatifs sur la diagonale,  $V$  est une matrice unitaire  $n \times n$  sur  $K$ , et  $V^*$  est la transposée conjuguée de  $V$ .

Dans la SVD, les entrées diagonales  $\sigma_i$  de  $S$  sont connues comme les valeurs singulières de  $M$ .

### 3.4.2.5 Tatouage audio par Substitution basé DCT

C'est une technique simple dans le domaine de la transformation [116]. Dans cette technique, les coefficients de transformation spécifiques du signal audio sont modifiés par un facteur d'échelle en fonction de la valeur de bit de la marque. Côté

extraction, la comparaison des coefficients modifiés est effectuée pour extraire la marque insérée. Certains coefficients ont les mêmes valeurs à différents endroits dans la gamme de fréquence de bande moyenne, et ils seront choisis pour effectuer la substitution car les coefficients de cette bande de fréquence sont moins affectés aux attaques et offrent une plus grande robustesse que les autres coefficients de DCT [116]. Le principal avantage de cette technique est qu'elle est une approche aveugle et offre plus de robustesse par rapport à la technique de substitution LSB.

Cette technique est un type de tatouage audio semi-fragile et offre une robustesse contre les attaques de bruit additif et de recadrage (*cropping*) mais elle a moins de transparence perceptuelle que le tatouage audio additif dans le domaine spatial.

#### 3.4.2.6 Tatouage Audio Multiplicatif basé DCT/DWT/SVD

Dans le domaine de transformation du tatouage, la marque est insérée dans les coefficients de transformation spécifiques du signal audio. Elle est généralement insérée dans des coefficients spécifiques de signal audio qui sont moins affectés en raison du bruit et fournissent la robustesse contre les attaques. Par conséquent, la marque image ou logo doit être mise à l'échelle en fonction de l'amplitude de ces coefficients. Ce type de tatouage est connu sous le nom de tatouage multiplicatif [116]. L'équation de base pour le tatouage multiplicatif est donnée ci-dessous :

$$T_s^* = T_s(1 + \alpha \times W) \quad (3.34)$$

Où  $T_s^*$  sont des coefficients modifiés du signal audio,  $T_s$  sont des coefficients originaux du signal audio,  $W$  est une marque et  $\alpha$  est un facteur de mise à l'échelle.

Dans cette technique, la sélection du facteur d'échelle joue un rôle important. Si la valeur du facteur de mise à l'échelle est élevée, alors plus de distorsion apparaît dans le signal audio tatoué, mais fournit une bonne image tatouée extraite. Il offre donc un bon compromis entre robustesse et transparence perceptuelle. Le tatouage multiplicatif pour le signal audio peut être mis en œuvre en utilisant trois transformations de base telles que DCT, DWT, et SVD. Le principal inconvénient de cette technique est qu'il s'agit d'une approche non aveugle. Mais cette technique offre une meilleure capacité de charge utile par rapport à la technique de substitution dans le domaine de transformation. Les résultats de simulation des techniques de tatouage audio multiplicatif en utilisant les transformées DCT et DWT ont montré qu'il s'agit d'un tatouage audio fragile [116].

#### 3.4.2.7 Tatouage audio additif basé SVD

Il s'agit d'une version modifiée du tatouage audio additif dans le domaine spatial. Dans cette technique, la marque est insérée dans les coefficients de transformation du signal audio. L'une des techniques la plus connue de tatouage audio additif dans le domaine de la transformation est la technique de décomposition en valeur singulière (SVD) [116]. Dans cette technique, la valeur singulière du signal audio est modifiée par le facteur de mise à l'échelle et la valeur singulière de la marque pour générer un signal audio tatoué. Le principal avantage de cette technique est que les matrices  $U$  et  $V$  de la marque sont nécessaires pour la récupération de la marque du côté de l'extraction. Les étapes pour l'intégration de marques sont expliquées ci-dessous.

1. Prendre le signal audio original et le convertir en matrice audio (M). Appliquer la SVD à la matrice audio (M) pour obtenir la valeur de la matrice U, la matrice S et la matrice V.
2. Prendre la marque, et appliquer SVD dessus pour obtenir sa matrice  $U_w$ , sa matrice  $S_w$  et sa matrice  $V_w$ .
3. Insérer la marque dans les valeurs de la matrice S du signal audio en utilisant l'équation ci-dessous :

$$S^* = S + \alpha \times S_w \quad (3.35)$$

Où  $S^*$  représente les valeurs singulières modifiées du signal audio, S pour les valeurs singulières originales du signal audio et  $S_w$  pour les valeurs singulières de la marque.

4. Appliquer la SVD inverse sur les valeurs modifiées de la matrice S avec la matrice U et la matrice V non modifiées pour obtenir la matrice audio tatouée
5. Convertir la matrice audio tatouée en vecteur pour obtenir le signal audio tatoué

Les étapes de l'extraction de la marque sont décrites ci-dessous :

1. Prendre le signal audio tatoué et le convertir en matrice audio tatouée (WM). Appliquer la SVD à la matrice audio tatouée (WM) pour obtenir la valeur de la matrice U, de la matrice S et de la matrice V.
2. Prendre le signal audio original et le convertir en matrice audio (M). Appliquer la SVD à la matrice audio (M) pour obtenir la valeur de la matrice U, de la matrice S et de la matrice V.
3. Prendre l'extraction de la valeur singulière de la marque à partir du signal audio tatoué en utilisant l'équation ci-dessous :

$$ES_w = \left(\frac{S^*}{S} - 1\right)/\alpha \quad (3.36)$$

Où  $ES_w$  sont les valeurs singulières recouvertes de la marque.

4. Appliquer la SVD inverse sur les valeurs extraites de la matrice Sw avec la matrice Uw non modifiée et la matrice Vw pour récupérer la marque tatouée.

Les résultats de simulation du tatouage audio multiplicatif basé SVD dans le domaine de transformation ont indiqué que la qualité de marque tatouée récupérée est supérieure à la technique de tatouage multiplicatif basée DCT et DWT. La mesure de performance a montré que cette technique est un type de tatouage audio robuste et qu'elle offre de meilleures performances par rapport aux techniques multiplicatives fondées sur la DCT et la DWT et aux techniques de tatouage audio du domaine spatial en termes de qualité de la marque récupérée [116] .

### 3.4.3 Tatouage audio de domaine Cepstral

Connu aussi comme domaine de log spectral, les données de cette méthode sont insérées dans des coefficients de cepstrum qui supportent bien les attaques de traitement de signal les plus courantes que les échantillons originaux dans le domaine

temporel. En plus, l'altération du cépstrum à des fréquences qui se trouvent dans les régions visiblement masquées de la majorité des trames du fichier audio de couverture assure l'inaudibilité du fichier audio résultant [77]. L'emploi de la modification du domaine cepstral a été proposé par Li et Yu, dans [117]. Dans ce schéma de tatouage, le signal de couverture est transformé en domaine cépstral et les données sont insérées dans le coefficient cépstrum sélectionné en manipulant la moyenne statistique de certains coefficients du cepstrum. Un modèle psychoacoustique intuitif a été utilisé pour contrôler l'audibilité de la distorsion introduite. Il a été démontré que la représentation de l'audio en cepstrum est très résistante à un grand nombre d'attaques, y compris la mise à l'échelle temporelle et la déformation de la hauteur de son les plus difficiles. Dans cette méthode, un taux d'insertion de 20 à 40 bps a été atteint tout en garantissant la robustesse aux attaques communes du signal et une transparence des données dissimulées. Gopalan dans [118], modifie légèrement les cépstrums de deux fréquences sélectionnées  $f_1$  et  $f_2$  dans chaque trame énergétique pour intégrer le bit '1' ou '0'. Cette méthode assurait un taux d'insertion fiable d'environ 54 bits/s. Pour plus de sécurité des données incorporées, Gopalan dans [119] suggère d'utiliser ce dernier algorithme et d'insérer des données avec des composantes différentes de fréquence arbitraires à chaque trame.

Liu et Lin dans [121] proposent une approche de tatouage audio basée sur le code BCH dans le domaine Cepstral. Cette technique tire parti de la caractéristique d'invariance de l'attaque dans ce domaine et de la capacité de correction des erreurs du code BCH pour accroître la robustesse du tatouage audio. En plus, le son tatoué présente une très haute qualité perceptive. Une technique de détection aveugle de la marque est développée pour identifier la marque intégrée sous différents types d'attaques. Les résultats de l'expérience montrent que la technique proposée présente une robustesse contre la plupart des attaques asynchrones par rapport aux techniques de tatouage audio existantes. Cependant au niveau transparence du schéma de tatouage, aucun test d'imperceptibilité n'a été réalisé.

#### **3.4.4 Tatouage audio de domaine hybride**

Ce type de tatouage audio est très utilisé dans la pratique. Dans ce type de technique, la combinaison de deux transformations de signal ou plus de deux est utilisée pour l'insertion de la marque. Dans cette technique, les coefficients hybrides du signal audio sont modifiés par la marque. Les combinaisons de transformations de signal les plus connues telles que DWT + DCT, DWT + SVD, et DWT + DCT + SVD sont utilisées dans le domaine hybride de tatouage audio [116].

Nematollahi et al. [122] ont proposé une technique de tatouage hybride utilisant la DWT et la SVD pour les signaux audio. Dans cette technique, le signal de parole est converti en trames. La DWT est appliqué à chaque trame pour obtenir des coefficients d'ondelette de trame. Les données de la marque sont insérées dans des valeurs singulières des coefficients d'ondelette approximatifs de chaque trame. Enfin, une SVD inverse et une DWT inverse sont appliqués à chaque trame modifiée pour obtenir le signal de parole tatoué.

Dhar et Shimamura [123] ont proposé deux techniques de tatouage hybride dans le domaine SVD pour le signal audio numérique. Ils ont proposé une technique basée sur DWT + DCT + SVD et une technique basée sur la transformée de Fourier rapide

(FFT) + SVD + Cartesian polar transform (CPT). Dans la première technique, les valeurs singulières des coefficients DCT d'approximation des coefficients d'ondelette du signal audio hôte sont modifiées en fonction des informations de la marque. Cette méthode fournit des sons tatoués hautement imperceptibles ainsi qu'une bonne robustesse contre diverses attaques. Dans la deuxième technique, les informations de la marque sont insérées dans les coefficients polaires cartésiens de valeurs singulières des coefficients FFT de basse fréquence du signal audio hôte. Les informations de la marque sont chiffrées par une technique de chiffrement chaotique. La charge de données de la méthode proposée est relativement plus élevée que celle des méthodes les plus récentes. En plus, elle montre une bonne robustesse contre diverses attaques.

Dey et al. [124] ont proposé une technique hybride de tatouage pour le signal utilisant DCT, DWT et SVD. Dans cette technique, les valeurs singulières des coefficients DCT des coefficients d'ondelette des coefficients de la marque en échelle de gris sont insérées dans les valeurs singulières des coefficients DCT des coefficients d'ondelette de la sous-bande HH du signal d'électrooculogramme.

Bhat et al. [125] ont proposé une technique de tatouage hybride pour le signal audio utilisant DWT et SVD. Dans cette technique, l'information de la marque est insérée dans les valeurs singulières des coefficients d'ondelette de basse fréquence du signal audio pour obtenir le signal audio tatoué.

Wang et al. dans [126] et Wang et Zhao dans [127] ont proposé une technique de tatouage pour le signal audio utilisant DCT et DWT. Dans cette technique, les coefficients DCT des coefficients d'ondelettes de basse fréquence sont modifiés en fonction des bits de la marque.

Les résultats d'expérimentation de la combinaison DWT+SVD [116] ont montré que la qualité de la marque récupérée est supérieure au tatouage audio du domaine spatial et à celui du domaine de transformation. La mesure des performances du tatouage audio hybride de DWT+ SVD dans le domaine de la transformation contre diverses attaques indique que cette technique est un type de tatouage audio robuste et offre de meilleures performances par rapport aux techniques de tatouage audio du domaine de transformation et du domaine spatial en termes de qualité de la marque récupérée.

Le tableau 3.3 résume les caractéristiques des méthodes de tatouage audio citées [116] :

Tableau 3.3 - Récapitulatif des caractéristiques des techniques citées

<i>N°</i>	<i>Technique</i>	<i>Domaine</i>	<i>Caractéristiques</i>
1	Tatouage Audio par Substitution LSB	Spatial	Facile à mettre en œuvre, Capacité élevée, moins de robustesse contre les attaques de tatouage. Fragile
2	Tatouage Audio Additif	Spatial	Fragile, utilisé pour l'authentification des droits d'auteur du signal hôte.
3	Tatouage Audio par Substitution DCT	Transformation	Tatouage semi-fragile, un peu robuste contre certaines attaques comme ajout de bruit et cropping, Imperceptibilité mauvaise
4	Tatouage Audio Multiplicatif basé DCT	Transformation	Qualité de l'image récupérée inférieur / DCT de substitution, Fragile



5	Tatouage Audio Multiplicatif basé DWT	Transformation	Qualité de l'image récupérée inférieure / DCT Multiplicatif, Tatouage fragile
6	Tatouage Audio Additif basé SVD	Transformation	Tatouage robuste, Qualité de l'image récupérée meilleure que le tatouage multiplicatif DCT et DWT et les techniques de tatouage du domaine spatial
7	Tatouage Audio de Domaine Hybride (DWT+SVD)	Hybride	Tatouage robuste, Qualité de la marque récupérée meilleure que toutes les autres techniques du domaine de transformation et du domaine spatial

### 3.5 TRAVAUX CONNEXES DANS LE DOMAINE DE TATOUAGE NUMERIQUE AUDIO

Dans cette section, nous passons en revue les différents travaux connexes récents que nous avons utilisés pour la comparaison de nos approches réalisées dans le cadre de cette thèse. Ces travaux de recherche sont présentés par domaine d'insertion. Les techniques de tatouage appliquées aux fichiers audio sont liées aux différents espaces ou domaines de représentation. À chaque domaine d'insertion correspondent différentes techniques de tatouage, Ce domaine peut être soit un domaine spatial, soit un domaine de transformation.

#### 3.5.1 Méthodes de tatouage audio dans le domaine spatial

Les méthodes spatiales permettent d'insérer la marque directement dans les échantillons de données et ne nécessitent aucune transformation. Ce sont des méthodes simples et peu coûteuses en termes de temps de calcul puisqu'elles ne nécessitent pas d'étape de traitement préalable. Elles sont dédiées au tatouage en temps réel requis dans des environnements à faible puissance.

- Hosny et al. [128] proposent un schéma de tatouage où deux échantillons successifs sont sélectionnés, puis, les bits de la marque sont insérés dans la 3ème et 4ème position des bits le moins significatif de l'échantillon sélectionné. Ensuite, un algorithme d'optimisation est utilisé pour minimiser l'erreur générée par le processus de dissimulation.
- Kundu et Kaur [129] proposent un schéma de tatouage où le message à cacher est chiffré par l'algorithme de cryptage carré de Vigenere. Ensuite, ces données sont incorporées dans les couches profondes par la méthode LSB (Least Significant Bit) modifiée. Par la suite, un chiffrement par transposition est utilisé pour chiffrer le fichier audio.
- Devi et Pugazhenthib [130] présentent une méthode de tatouage (Figure 3.18) pour insérer des données dans un fichier audio en utilisant la technique d'insertion LSB. Le fichier audio est converti en un flux binaire. Le premier bit du message à intégrer est inséré dans le bit le moins significatif du premier échantillon. Le processus est répété jusqu'à ce que les bits du message soient entièrement insérés dans le fichier audio.

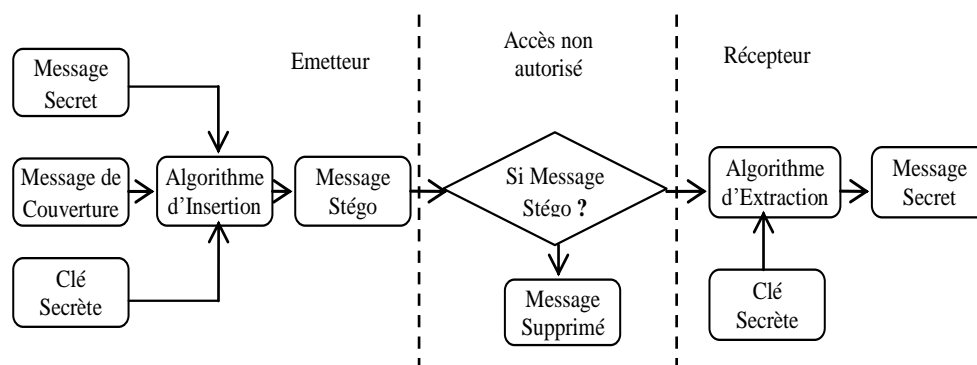


Fig. 3.18 - Schéma d'insertion de Devî et Pugazhenthîb [130]

### 3.5.2 Méthodes de tatouage audio dans le domaine de transformation DFT / DCT

Les domaines de transformation ont été largement étudiés dans le contexte du codage et de la compression, et de nombreux résultats de recherche peuvent être appliqués au tatouage numérique [131]. L'utilisation d'un domaine de transformation spécifique, tel que la DCT ou la DWT, a deux objectifs. Elle décorrelle les valeurs de l'échantillon original et concentre l'énergie du signal original dans quelques coefficients seulement [132]. De nombreux travaux ont été proposés pour le tatouage audio basé sur la transformée de Fourier.

- Abdelwahab et al. [133] proposent un schéma de tatouage audio qui combine la SVD (décomposition en valeur singulière) et la FRT (transformée de Fourier fractionnée). Dans cette approche, le signal est transformé en une matrice 2D, puis une SVD est appliquée à cette matrice (S). La matrice W (image de la marque) est ajoutée à la matrice S. Une SVD est appliquée à la matrice résultante, qui est ensuite transformée en une matrice 1D.
- Ortiz et al. [134] ont proposé une approche aveugle et robuste pour le tatouage audio. Dans cette approche, le signal audio original est d'abord divisé en segments. Une FFT est ensuite effectuée sur chaque segment. Les coefficients consécutifs de moyenne fréquence de chaque segment sont sélectionnés comme ensemble de données pour l'insertion de la marque. L'élément ayant la plus grande amplitude est sélectionné pour intégrer un bit de la marque.
- Dans la méthode de tatouage proposée par Hu & Lee [135], les segments audio appropriés pour la dissimulation sont d'abord sélectionnés. Chaque segment est ensuite divisé en trames (sans chevauchement). Ensuite, une FFT est appliquée à chaque trame. Les bits de la marque sont alors intégrés dans les régions de basse fréquence.
- Seyed et al. [136] proposent une nouvelle approche pour le tatouage audio basée sur les valeurs des niveaux d'énergie dans le domaine FFT. Cette méthode emploie une procédure d'incorporation régulière sur la base de la fermeture de l'énergie des sous-trames à la séquence mathématique Lucas, ce qui rend possible l'intégration de deux bits dans chaque trame audio.
- Jayarani et al. [137] ont proposé un schéma de tatouage nul ou zéro pour le tatouage des fichiers audio. L'algorithme est basé sur la transformée de Fourier à temps court (Short Time Fourier Transform - STFT). La marque est générée à l'aide du spectre de phase STFT, ce qui préserve l'originalité du fichier. La marque générée est ensuite stockée dans une base de données pour vérifier si le signal est altéré ou non.

De nombreux travaux basés sur des transformées comme DCT, DWT et SVD ont été aussi proposés pour le tatouage audio :

- Dans l'approche proposée par Irawati et al. [138], en utilisant un schéma de tatouage audio basé sur la modulation d'indice de quantification (QIM), une LWT (Lifting Wavelet Transform) est appliquée, puis une sous-bande est sélectionnée et transformée du domaine temporel au domaine fréquentiel par DCT (Figure 3.19). Chaque trame des coefficients DCT est décomposée en une matrice orthogonale et une matrice triangulaire par QR. En utilisant l'encodage QIM, la marque est ensuite insérée dans une matrice triangulaire en quantifiant les données à une valeur correspondant au quantificateur contenu dans la marque.

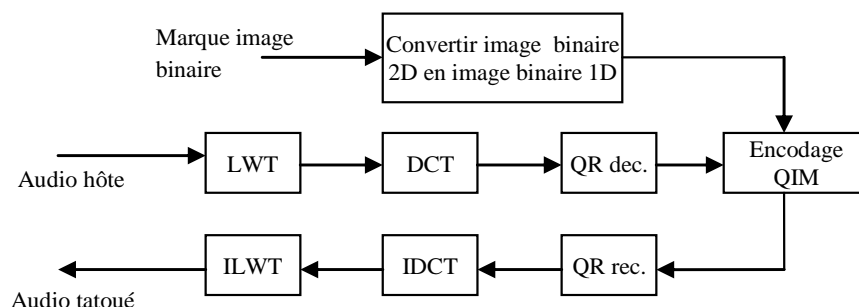


Fig. 3.19 - Schéma d'insertion d'Irawati et al. [138]

- Karajeh et Maqableh [139] proposent un système de tatouage audio combinant la transformation en cosinus discrète (DCT) et la décomposition de Schur. Une DCT est d'abord appliquée au signal audio, puis la décomposition de Schur est appliquée à la bande de fréquence moyenne des coefficients de la DCT. Les bits de la marque seront substitués aux LSB des éléments diagonaux de la matrice triangulaire résultante  $S$ . L'utilisation de la DCT augmente la robustesse et la décomposition de Schur améliore la transparence du processus d'intégration.
- Xue et al. [140] proposent un schéma de tatouage combinant la décomposition différentielle en valeur singulière (SVD) et la segmentation de parité. Une DCT est appliquée, puis les coefficients résultants sont divisés en deux segments en fonction de l'ordre de parité. L'énergie des segments appariés sera approximativement égale, ce qui réduira les changements dans les valeurs singulières pendant le processus de dissimulation. Cela permettra d'atteindre un haut degré d'imperceptibilité.

### 3.5.3 Méthodes de tatouage audio dans le domaine multi résolution

#### DWT

De nombreux travaux basés sur la transformée DWT du domaine de multi résolution ont été aussi proposés :

- Jiang et al. [141] proposent une nouvelle méthode de tatouage audio dans laquelle la marque est générée en utilisant le système de chaos et ensuite chiffrée (Figure 3.20), les trames sont ensuite divisées de manière adaptative. Cette division est basée sur les caractéristiques globales de la trame et est effectuée par synchronisation directe en utilisant un numéro de séquence de trame correspondant.

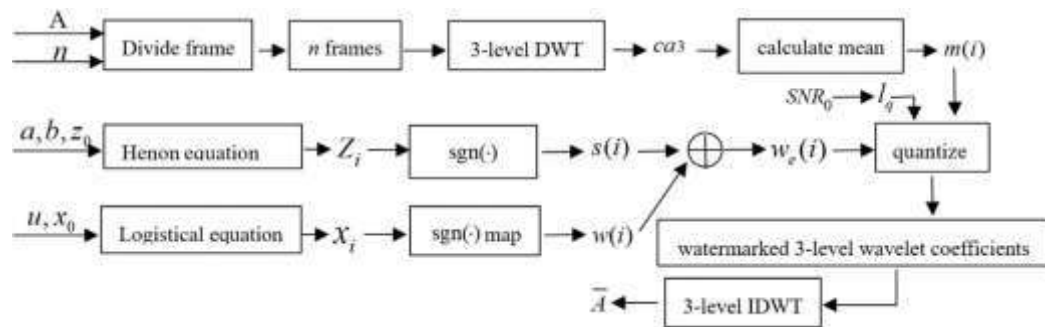


Fig. 3.20 - Schéma d'insertion de Jiang et al. [141]

- Saadi et al. [142] proposent un schéma de tatouage qui combine l'utilisation de la DCT et de la DWT. Après avoir cadré le signal, une DWT est appliquée, puis une DCT est appliquée à chaque trame. Chaque trame est divisée en deux segments par sous-échantillonnage pour obtenir une bonne corrélation. Pour augmenter la sécurité de la marque, la transformation d'Arnold est appliquée au message à intégrer.
- Kaur et Dutta [143] proposent un algorithme de tatouage qui se concentre sur la dissimulation des données en utilisant la transformation SVD du signal audio dans le domaine DWT. Chaque bit de la marque est caché dans le signal audio en utilisant la quantification moyenne pour permettre l'insertion d'une grande quantité d'informations tout en maintenant une bonne qualité audible du signal.
- Hu et al. [144] proposent une nouvelle méthode de tatouage audio dans laquelle une intégration binaire est effectuée dans la sous-bande d'approximation basse fréquence. Une transformée en ondelettes ascendantes (LWT - Lifting Wavelet Transform) à trois niveaux est d'abord appliquée. Ensuite, la séquence triée des quantités de coefficients d'approximation est remodelée en fonction du débit binaire prédit du tatouage.

### 3.6 CONCLUSION

Dans ce chapitre, nous avons présenté le contexte technique qui entoure le tatouage audio numérique et les différents travaux dans la littérature qui ont fait avancer la technologie dans ce domaine, nous avons détaillé le tatouage audio numérique en présentant l'architecture générale du système avec ses propriétés et conditions requises, nous avons ensuite dressé une taxonomie des techniques de tatouage audio classé selon différents critères : le domaine d'insertion, la perception humaine, la technique d'insertion, le champ d'application et la méthode de détection et d'extraction. En plus, nous avons présenté les différentes techniques de tatouage audio des différents domaines d'insertion avec une analyse comparative entre ces techniques ainsi que leurs avantages et leurs inconvénients. En fin nous avons passé en revue les différents travaux connexes et récents que nous avons utilisés pour la comparaison de nos approches réalisées dans le cadre de cette thèse.

Le tatouage numérique est devenu un domaine de recherche majeur pour la protection du contenu multimédia en raison des progrès rapides des technologies multimédia, de l'Internet à haut débit et des installations de communication, de la disponibilité bon marché des dispositifs d'affichage, etc. L'insertion d'une marque dans un signal audio sans en altérer la qualité est plus difficile que pour les images et les vidéos. Plusieurs schémas d'insertion de marque existent dans la littérature. Ils sont basés soit sur le domaine temporel, soit sur le domaine de la transformation, soit ils utilisent des techniques combinées des deux domaines.

Dans ce chapitre, plusieurs techniques de tatouage utilisées dans les travaux récents des chercheurs ont été présentées. Le tatouage audio, en particulier, aborde les questions liées à la nécessité de sécuriser et de préserver l'intégrité des données cachées dans les communications vocales et audio, même lorsque celles-ci passent par des canaux non sécurisés. Nous avons donné une revue de l'état actuel de la littérature dans les techniques et les approches de tatouage audio numérique. Nous avons présenté quelques-unes des techniques de tatouage audio les plus intéressantes. Nous avons discuté de leur potentiel et de leurs limites pour assurer une communication sécurisée. De notre point de vue, une comparaison et une évaluation des techniques examinées ont également été données. L'avantage d'utiliser une technique par rapport à une autre dépend fortement du type d'application et de ses exigences telles que la capacité de dissimulation ou le type d'attaques qui pourrait rencontrer le signal transmis.

L'une des exigences essentielles d'un système de tatouage audio numérique est de présenter une robustesse, une imperceptibilité, et une capacité élevées. Cependant, aucun algorithme ne peut à lui seul répondre aux exigences d'un tel système, car ces paramètres se compensent mutuellement. Par conséquent, un compromis est toujours nécessaire pour choisir un schéma particulier pour une application spécifique. La robustesse, l'imperceptibilité et la capacité sont considérées comme des critères de performance dans l'examen des systèmes de tatouage audio numérique dans les domaines temporel et de transformation. Le quatrième chapitre de cette thèse traite cette problématique, il sera donc consacré à l'étude des critères et des mesures utilisés pour évaluer la performance de ces systèmes.

## **4. Critères de Mesure de Performance des Systèmes de Tatouage Audio Numérique**

---

### **Sommaire**

---

4.1	<b>Introduction</b> .....	81
4.2	<b>Le système auditif humain</b> .....	81
4.3	<b>Les attaques sur les systèmes de tatouage audio numérique</b> .....	88
4.4	<b>Evaluation de l'imperceptibilité</b> .....	102
4.5	<b>Evaluation de la robustesse</b> .....	112
4.6	<b>Evaluation de la fiabilité</b> .....	115
4.7	<b>Evaluation de la capacité</b> .....	116
4.8	<b>Conclusion</b> .....	117

---

## 4.1 INTRODUCTION

L'évaluation d'un schéma de tatouage audio est liée directement aux propriétés de base requises au système, que sont l'imperceptibilité du tatouage, la robustesse contre les attaques, la fiabilité et la capacité d'insertion. Plusieurs essais en plusieurs points de la chaîne de communication, à savoir l'insertion, la détection et le canal de transmission, doivent être testés.

Dans ce chapitre nous commençons par présenter les propriétés du système auditif humain (HAS), car Il est important de comprendre comment le HAS perçoit le son pour le développement d'un système de tatouage audio efficace, ensuite nous présentons les groupes d'attaques qui sont à l'origine des opérations courantes de traitement de signal appliquées sur les fichiers audio tatoués durant leur utilisation ou leur diffusion sur internet, nous faisons aussi un état des outils de benchmark cités dans la littérature pour tester l'imperceptibilité et la robustesse des systèmes de tatouage audio face à ces attaques, et en fin nous faisons état de différentes mesures proposées par la littérature et nous présentons les méthodes et les métriques d'évaluation des schémas de tatouage ainsi que les différents types de tests et de mesures utilisés pour l'évaluation des schémas de tatouage audio. Certains de ces mesures comme les tests d'imperceptibilité et de robustesse ont été utilisés dans l'évaluation de nos approches de tatouage audio proposées.

## 4.2 LE SYSTEME AUDITIF HUMAIN

La modélisation psycho-acoustique est importante dans le codage et le tatouage audio pour garantir que les modifications du signal original restent imperceptibles [145]. Comparé au système visuel humain (*Human Visual System* - HVS), le système auditif humain HAS (*Human Auditory System*) est beaucoup plus sensible, ce qui rend le tatouage audio plus difficile que celui de l'image (Cox et al., 2002) [146]. Le HAS peut détecter un signal dont la gamme de fréquences est supérieure à  $10^3:1$  et dont la puissance est supérieure à  $10^9:1$  (Painter et al., 2000) [147]. Il est important de comprendre comment le HAS perçoit le son pour le développement d'un système de tatouage audio efficace.

Même s'il n'existe pas encore de modèles exacts de HAS, de grands progrès ont été réalisés dans le domaine de la psycho-acoustique pour modéliser la perception auditive humaine ainsi que les capacités d'analyse temps-fréquence de l'oreille interne. La principale propriété de la perception audio réside dans le phénomène de masquage, qui comprend le pré-masquage et le post-masquage. Un autre problème de la perception audio est le seuil absolu d'audition [145].

### 4.2.1 Propriétés de l'audition humaine monaurale

Les principes du système auditif humain et les propriétés physiologiques sous-jacentes de l'oreille ont été étudiés par des chercheurs dans les domaines de la médecine, de la biologie et de l'ingénierie. Des études psycho-acoustiques complètes ont été réalisées et décrites par plusieurs chercheurs comme Fastl et Zwicker [148] et Moore [149], entre autres.

De nombreux effets psycho-acoustiques peuvent être expliqués par les propriétés physiologiques du processus auditif lui-même. L'oreille externe et le canal auditif recueillent les ondes de pression de l'air qui sont transmises sous forme de vibrations mécaniques via le tympan et l'oreille interne à la cochlée remplie de liquide. La forme de la cochlée ressemble à une spirale effilée semblable à un escargot enroulé (son nom est dérivé du mot grec ancien signifiant "coquille d'escargot"). La surface interne de la cochlée est recouverte de cellules sensorielles composées de petits poils reliés à des terminaisons nerveuses. En fonction de la fréquence du son entrant, différentes zones de la cochlée résonnent, ce qui fait vibrer les cellules ciliées de la région concernée ("organisation tonotopique") [45]. À leur tour, les terminaisons nerveuses stimulées envoient des impulsions au cerveau, ce qui lui permet de percevoir et de distinguer des sons de fréquence de niveau et de tonalité différente, La figure 4.1 donne une illustration de l'oreille humaine.

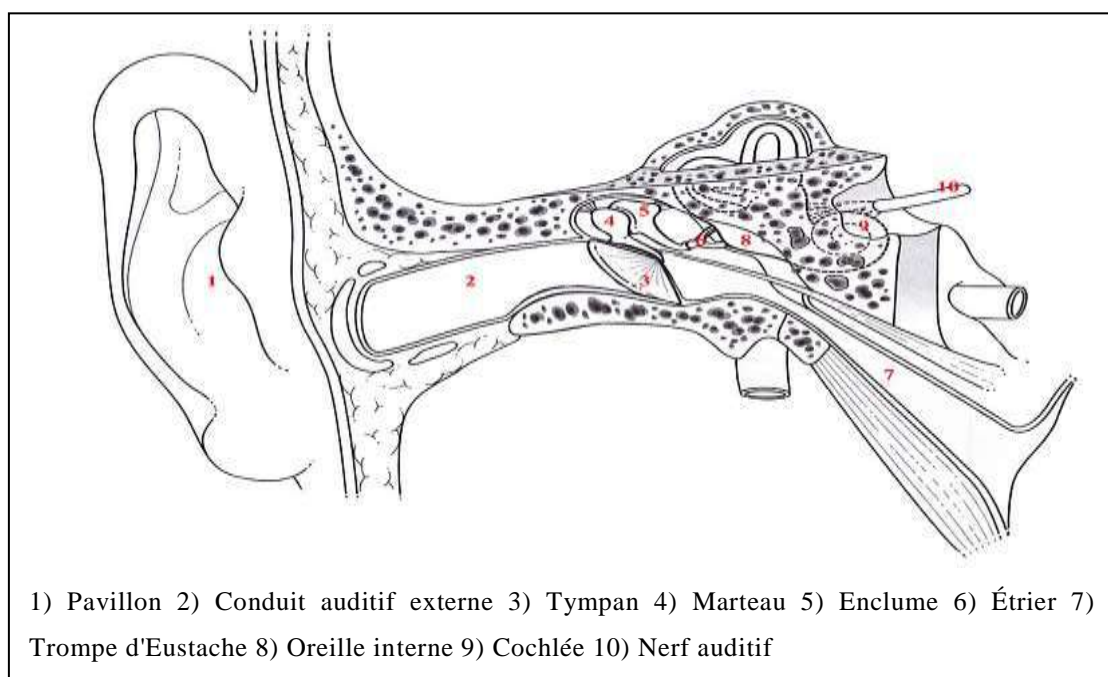


Fig. 4.1 - Diagramme de l'oreille (Source Wikipédia).

En raison de ces propriétés physiologiques du processus auditif, la perception et la résolution de la fréquence ne sont pas linéaires et deviennent plus grossières aux fréquences élevées. La largeur de bande des filtres auditifs du HAS, appelée "*largeur de bande critique*", décrit la largeur de bande dans laquelle les composants du signal interagissent perceptivement les uns avec les autres au lieu d'être perçus comme indépendants les uns des autres. La modélisation de cette largeur de bande des filtres auditifs permet d'établir des échelles de fréquence motivées par la perception, comme l'échelle de Bark proposée par Zwicker [148], [150] ou l'échelle de la largeur de bande rectangulaire équivalente (ERB) de Moore et Glasberg [151], [152] (voir la figure 4.2).



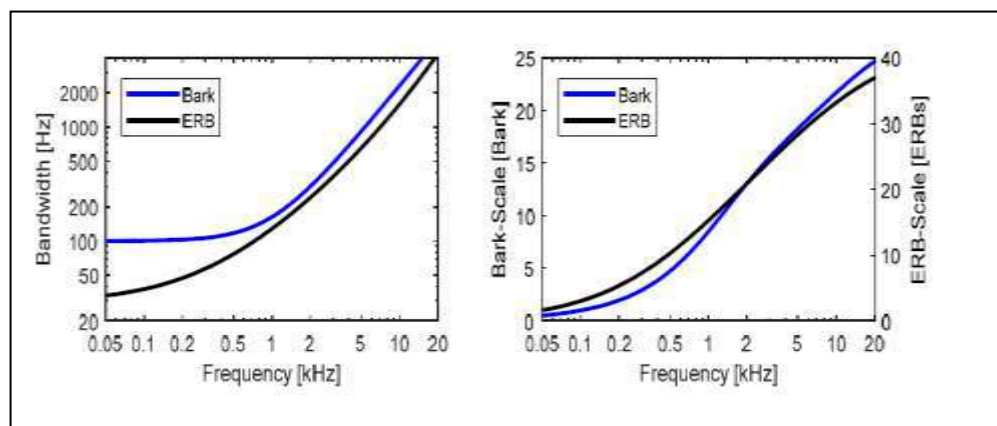


Fig. 4.2 - Largeur de bande critique et échelles perceptives résultantes, telles que modélisées par l'échelle de Bark et la largeur de bande rectangulaire équivalente (ERB) [45].

#### 4.2.2 Le seuil absolu d'audition (ATH -Absolute Threshold of Hearing)

En psycho acoustique, l'intensité d'un son est mesurée en termes de niveau de pression acoustique SPL (*Sound Pressure Level*). Le SPL donne le niveau de pression acoustique en décibels (dB) par rapport au niveau de référence défini au niveau international  $p_0 = \mu\text{Pa}$ , ou  $2 \times 10^{-5} \text{ N/m}^2$  (Newton par mètre carré), qui correspond au seuil d'audibilité à 1000 Hz. Ce niveau est donné par

$$\text{SPL} = 20\log_{10}\left(\frac{p}{p_0}\right) \text{ dB} \quad (4.1)$$

Où SPL est le niveau de la pression acoustique mesuré,  $p$  est le niveau de la pression acoustique du stimulus atteignant l'oreille, en Pascals [Newton par mètre carré ( $\frac{\text{N}}{\text{m}^2}$ )] (Spanias, et al. 2007)[153].

Bien que le HAS soit très sensible par rapport à l'HVS (*Human Visual System*), il a ses propres limites. Le HAS ne peut pas percevoir le son si son SPL est inférieur à un certain seuil. Ce seuil est appelé seuil absolu d'audition, qui détermine l'énergie d'un son pur qui peut être détecté par le HAS dans un environnement sans bruit. Le seuil absolu d'audition est une fonction non linéaire qui varie en fonction de la fréquence audio et elle est représentée dans la Figure 4.3 (Painter et al., 2000) [147].

Dans (Zwicker et al., 1990)[148], nous trouvons que la HAS est plus sensible aux sons dont la gamme de fréquences se situe entre 2 kHz et 5 kHz et qu'elle n'est pas sensible aux sons dont la fréquence est supérieure à 10 kHz. La zone située sous le seuil absolu d'audition est appelée zone de silence et le signal audio qui tombe dans la zone de silence n'est pas perceptible [145].

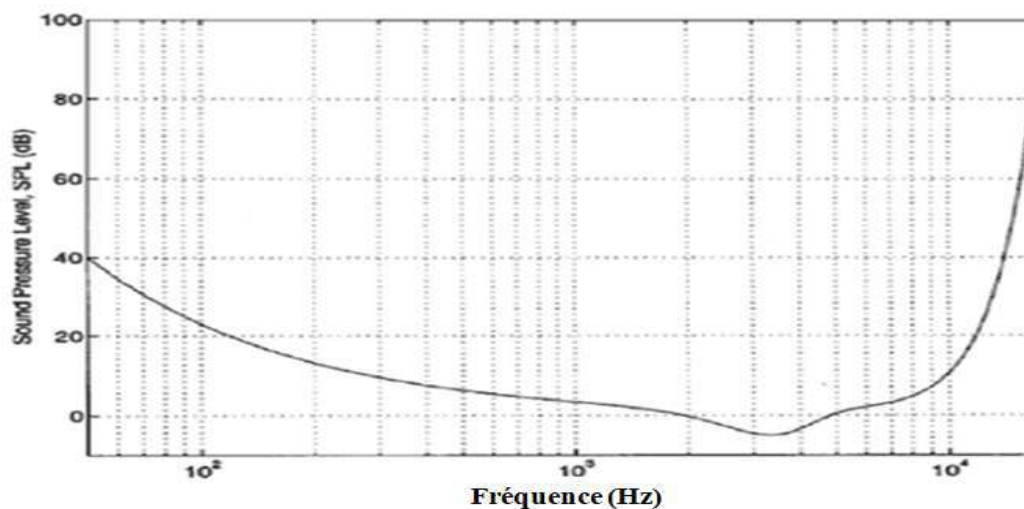


Fig. 4.3 - Le seuil absolu d'audition.

### 4.2.3 Les bandes critiques

Des expériences ont montré que la sensibilité humaine aux événements acoustiques est liée à l'échelle de fréquence irrégulière. Le terme "*bande critique*" décrit des régions de sensibilité équivalente dans cet échelle de fréquence et est définie comme la bande de fréquence à l'intérieur de laquelle l'intensité sonore d'une bande de son distribuée en continu et de SPL constant est indépendante de sa bande passante (Atal et al., 1984) [154]. La bande critique est évaluée sur l'échelle dite de *Bark*. Comme les bandes critiques sont espacées de façon irrégulière (voir le tableau 4.1), l'échelle "*Bark*" est une échelle de fréquence non linéaire. Il ya une correspondance entre les fréquences par seconde (Hz) et l'échelle de Bark.

L'oreille divise le spectre audible en 24 bandes critiques (également appelées "*filtres cochléaires*") adjacentes que l'on peut comparer à des filtres passe-bande 1/3 d'octave [155].

Tableau 4.1 - Répartition fréquentielle des 24 bandes critiques de l'oreille humaine

Bande critique	Fréquence (Hz)			Bande critique	Fréquence (Hz)		
	Basse	Haute	Largeur		Basse	Haute	Largeur
0	0	100	100	13	2000	2320	320
1	100	100	100	14	2320	2700	380
2	200	100	100	15	2700	3150	450
3	300	100	100	16	3150	3700	550
4	400	510	110	17	3700	4400	700
5	510	630	120	18	4400	5300	900
6	630	770	140	19	5300	6400	1100
7	770	920	150	20	6400	7700	1300
8	920	1080	160	21	7700	9500	1800
9	1080	1270	190	22	9500	12000	2500
10	1270	1480	210	23	12000	15500	3500
11	1480	1720	240	24	15500	22050	6550
12	1720	2000	280				

Il est à noter que la largeur des bandes de 0 à 3 est constante. A partir de la bande critique n°4, la largeur des bandes augmente logarithmiquement. On note aussi que plus la fréquence est élevée, plus la bande critique s'élargit. La figure 4.4 montre les bandes critiques et leur largeur de bande à chaque fréquence centrale (Zwicker et al., 1991) [156].

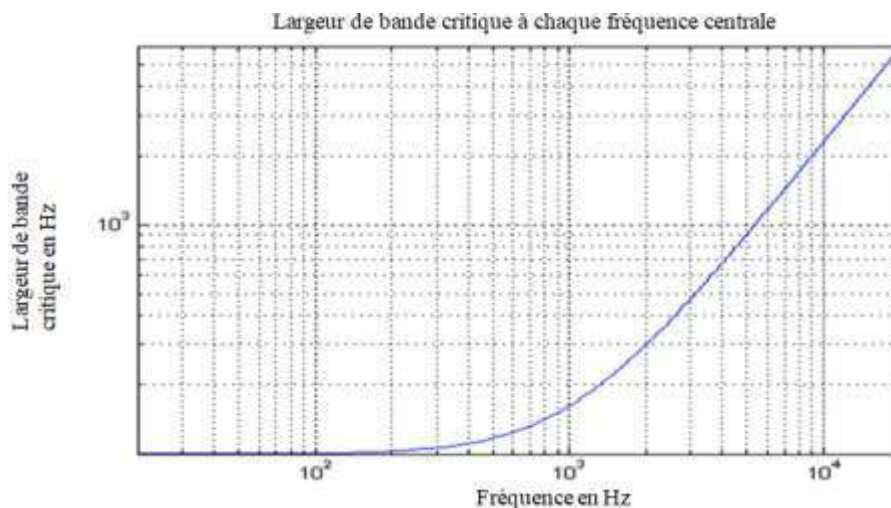


Fig. 4.4 - Largeur de bande critique à chaque fréquence centrale

#### 4.2.4 Le phénomène de masquage

Le masquage auditif désigne le phénomène par lequel un son devient inaudible en raison de l'existence d'un autre son. Le son masqué est appelé masqué « *masked* » et le son qui masque l'autre son est appelé masquant « *masker* ». Il existe deux types de phénomènes de masquage auditif : le masquage simultané et le masquage non simultané, qu'on les appelle également le masquage fréquentiel et le masquage temporel [145].

La distribution standard des bandes critiques (tableau 3.1) se produit lorsque deux ou plusieurs sons sont présents en même temps et que le signal le plus faible est rendu imperceptible en raison de la présence du signal le plus fort, en d'autres termes, le signal faible est masqué par le signal le plus fort. Chaque fois qu'un stimulus est présent, il crée un seuil de masquage et rend inaudible tout signal qui se situe en dessous de la courbe de masquage.

Un exemple courant de masquage est la façon dont le bruit d'un oiseau (sonde) peut être noyé par le bruit d'une voiture qui passe (masquant) dans la perception auditive humaine. Une catégorie de ces effets de masquage est connue sous le nom d'"*effets de masquage spectral*" ou de "*masquage simultané*" (Zwicker et al., 1990)[148].

Dans la figure 4.5, La ligne brisée représente le seuil absolu d'audition, la ligne solide illustre le seuil de masquage du à la présence d'un signal masquant (par exemple, bruit à bande étroite), en raison duquel les signaux plus faibles aux fréquences voisines deviennent inaudibles.

Nous remarquons à partir de la figure 4.5, les aspects suivants [45] :

- Un seuil d'audition du silence dépendant de la fréquence décrit le niveau de pression acoustique (SPL) minimal d'un son pour être perceptible de manière isolée et dans des conditions extrêmement calmes.
- En présence d'un masquant, la courbe du seuil du silence se transforme en un seuil de masquage, qui présente une augmentation en forme de cloche des fréquences à proximité du masquant, en fonction de sa fréquence, de son niveau et du type de signal. Tout son en dessous de ce seuil est masqué par le signal le plus fort, et donc inaudible pour l'auditeur moyen.
- Les effets de masquage sont les plus forts pour les signaux qui se trouvent dans la bande passante critique du masquant. À l'intérieur de la largeur de bande critique, le seuil de masquage reste constant. En plus, les effets de masquage s'étendent aux fréquences situées au-delà de la largeur de bande critique (masquage inter-bande). La pente supérieure du seuil de masquage dépend de plusieurs facteurs, tels que la fréquence absolue et le niveau de pression acoustique du masquant, tandis que la pente inférieure ne présente pratiquement aucune dépendance vis-à-vis du niveau.
- L'intensité de l'effet de masquage varie selon le type de masquant, c'est-à-dire le son ou le bruit (à bande étroite). L'effet de masquage varie, alors que les "masquant" de type bruit peuvent très bien masquer les signaux de type tonalité (jusqu'à environ 6 dB), les "masquant" de type sonore ne peuvent masquer le bruit que dans une mesure beaucoup plus faible (environ 20 dB) (Hellman, 1972) [157].

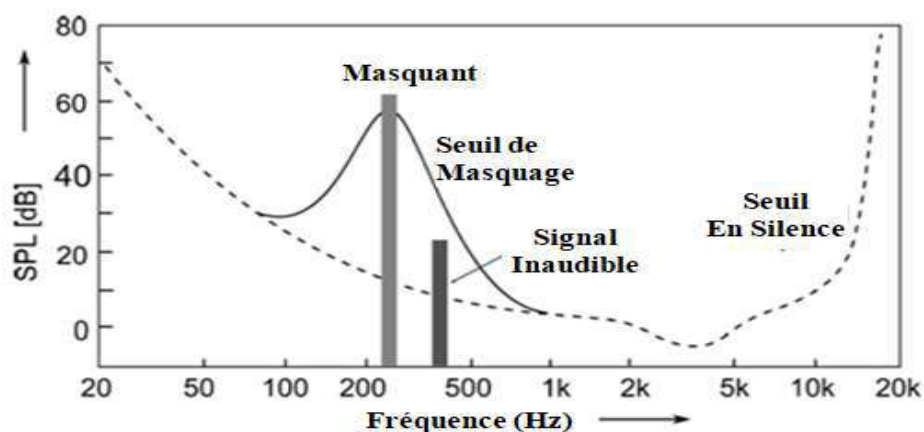


Fig. 4.5 - Illustrations des effets de masquage spectrale.

Il existe de nombreux types de masquage simultané et les trois principaux types simplifiés de masquage simultané sont *bruit-masquage-tonalité* (*noise-masking-tone-NMT*) (Scharf, 1970) [158], *tonalité-masquage-bruit* (*tone-masking-noise-TMN*) (Hellman, 1972) [157], et *bruit-masquage-bruit* (*noise-masking-noise-NMN*) (Akansu, et al. 1996) [159].

Dans le scénario *bruit-masquage-tonalité*, un bruit à bande étroite avec une largeur de bande de 1 Bark masque la tonalité dans la même bande critique, lorsque l'intensité de la tonalité est inférieure à certains dB de l'intensité du bruit masquant. La différence d'intensité entre le masquant (bruit) et le masqué (tonalité) est appelée le rapport de masquage du signal SMR (*signal masking ratio*) qui est le plus faible lorsque la fréquence du son est proche de la fréquence centrale du bruit. En général, le SMR pour la tonalité masquant le bruit est d'environ 4 à 5 dB, ce qui signifie que lorsque le signal sonore a une intensité inférieure de 4 ou 5 dB à l'intensité du bruit

dans la même bande critique, la tonalité devient inaudible. Un exemple est illustré à la figure 4.6 où une tonalité pure de 410 Hz avec un SPL (niveau de pression sonore) de 76 dB est rendue inaudible par le bruit à bande limitée centré sur 410 Hz, avec un SPL de 80 dB (Spanias, et al. 2007)[153].

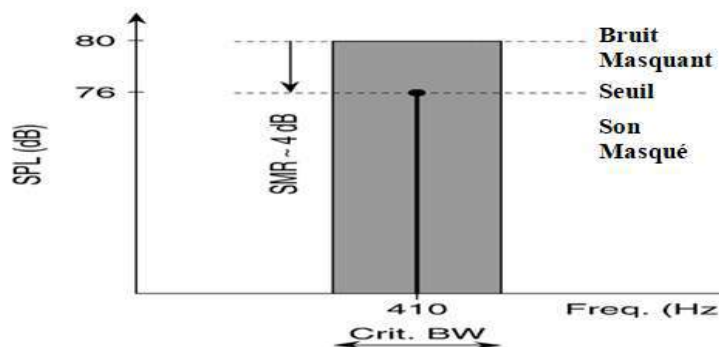


Fig. 4.6 - Exemple de bruit masquant tonalité.

Par contre, le cas *tonalité-masquage-bruit*, est le phénomène où le signal de bruit devient inaudible ou masqué par l'excitation d'un signal sonore dans la même bande critique. Des études ont montré que le SMR minimal général pour qu'une tonalité puisse masquer un bruit est d'environ 21 à 28 dB. La figure 4.7 montre un exemple de cas d'une tonalité masque un bruit où le bruit à bande étroite centré à 1 kHz est masqué par une tonalité pure à 1 kHz, à noter toutefois que l'intensité du bruit est inférieure de 24 dB à celle de la tonalité (Spanias, et al. 2007) [153].

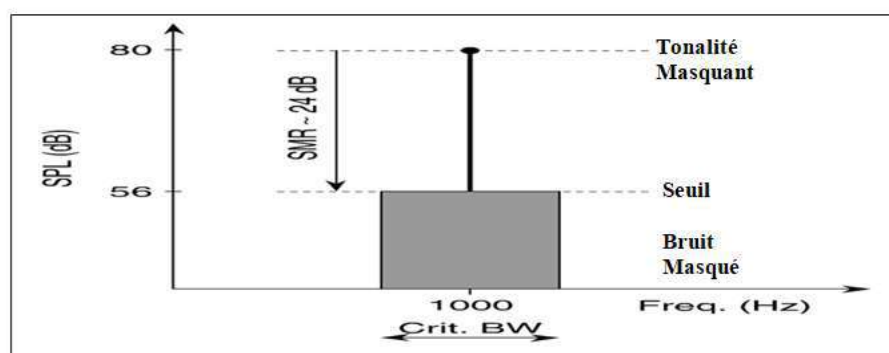


Fig. 4.7 - Exemple de tonalité masquant bruit.

Dans le scénario du *bruit-masquage-bruit*, un bruit à bande étroite est masqué par un autre bruit à bande étroite. Une étude a montré que les bruits à large bande peuvent produire environ 26 dB dans le cas du bruit masquant un bruit[145]. Le bruit est beaucoup plus efficace que le signal sonore comme masquant, créant un SMR beaucoup plus élevé que le signal sonore. Il est facile pour un bruit à bande étroite de rendre inaudible un signal sonore dans la même bande critique avec un SMR d'à peine 4 dB. Cependant, pour masquer le bruit à bande étroite dans la même bande critique, un signal sonore doit avoir une intensité de 21 à 28 dB supérieure à celle de signal du bruit.

Un signal d'excitation peut non seulement créer un effet de masquage simultané et rendre inaudibles d'autres signaux audio plus faibles dans la même bande critique, mais il peut également créer un tel effet de masquage sur des bandes presque critiques. Ce phénomène est appelé la propagation du masquage. Cet effet est souvent modélisé approximativement par une fonction d'étalement triangulaire avec des pentes de 25 et 10 dB par Bark (Painter et al., 2000)[147].

#### 4.2.5 Le masquage temporel

Un stimulus fort peut non seulement générer un masquage simultané, mais aussi créer des effets de masquage avant ou après sa présence. Un tel phénomène de masquage est appelé masquage non-simultané ou masquage temporel. Le phénomène de masquage avant l'apparition d'un son est appelé pré-masquage et celui après sa présence est appelé post-masquage [145].

La figure 4.8 (Painter et al., 2000) [147] illustre la propriété de masquage non-simultané du HAS. Il faut noter que si le pré-masquage ne dure que quelques millisecondes, le post-masquage peut durer jusqu'à 200 millisecondes.

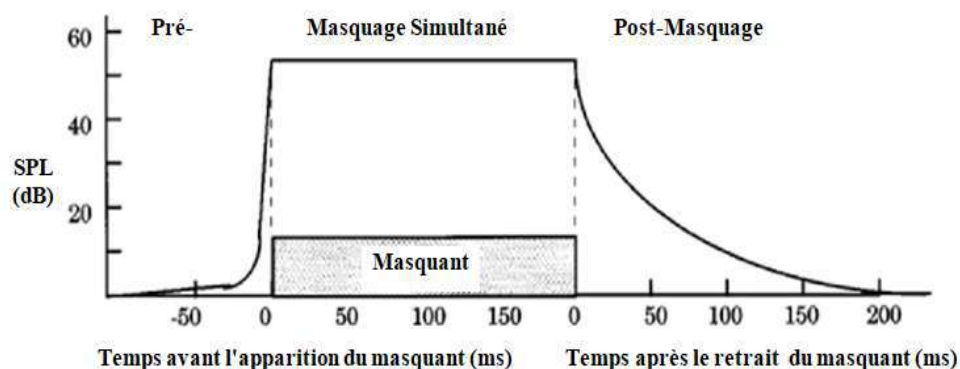


Fig. 4.8 - Propriété de masquage temporel du HAS.

### 4.3 LES ATTAQUES SUR LES SYSTEMES DE TATOUAGE AUDIO NUMERIQUE

Un système générique de tatouage audio numérique est représenté sur la figure 4.9. Les rectangles pleins spécifient les phases principales de manipulation du signal dans un système de tatouage audio et les termes dans les rectangles en pointillés représentent les utilisateurs possibles qui manipulent les données [68].

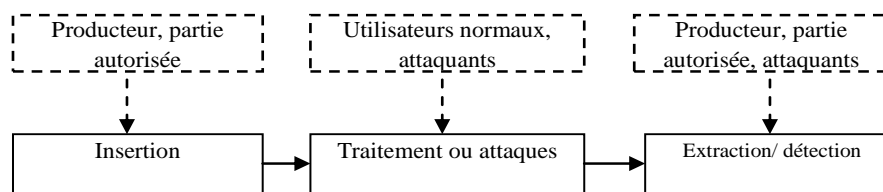


Fig. 4.9 - Schéma fonctionnel générique d'un système de tatouage audio numérique

L'efficacité d'un système de tatouage audio est caractérisée par plusieurs critères de performance [153], à savoir l'imperceptibilité, la robustesse, la sécurité, la capacité et la complexité de calcul, etc. L'imperceptibilité caractérise la fidélité des données audio tatouées, indiquant que les marques insérées doivent introduire des changements perceptiblement indiscernables dans le signal hôte. Par conséquent, les termes "fidélité", "transparence" et "inaudibilité" sont parfois utilisés de manière équivalente à l'imperceptibilité. La robustesse fait référence à la possibilité de réussir l'extraction de la marque lorsque le signal tatoué a été attaqué intentionnellement ou non. Il s'agit de la caractéristique la plus complexe pour un système de tatouage audio en raison de la variété des attaques. La sécurité signifie que le système doit être conçu de manière à ce que seules les parties autorisées soient en mesure d'extraire les marques insérées. La quantité d'informations qui peut être intégrée dans des données hôtes est appelée capacité. Enfin, le système conçu doit être efficace sur le plan de calcul informatique.

Pour des raisons d'efficacité, la plupart des travaux de recherche prennent principalement en compte les critères d'imperceptibilité et de robustesse parmi les critères [68], car elles déterminent la performance clé de la plupart des systèmes de tatouage audio existants. En revanche, la sécurité est généralement obtenue par l'utilisation de clés aléatoires cryptographiques, qui est largement utilisée dans la plupart des solutions existantes.

Il faut noter que les utilisateurs normaux peuvent "attaquer" involontairement le signal tatoué pendant les processus de compression avec perte, d'égalisation ou d'ajout d'effets, etc. Ainsi, nous appelons ces "attaques de traitement" d'attaques non intentionnelles, tandis que les attaques délibérées visant à détruire ou à supprimer les marques sont appelées attaques intentionnelles [68].

Bien que, généralement, il existe un compromis entre l'imperceptibilité et la robustesse, parfois, pour assurer la validité du système, l'imperceptibilité peut être compromise pour obtenir une meilleure robustesse [68], surtout lorsqu'il s'agit de faire face à des attaques avancées. Par exemple, pour traiter efficacement les attaques de désynchronisation, les valeurs ODG sont fixées à environ -1 dans [160] et [161], indiquant que les marques sont réellement perceptibles mais non gênantes. En raison de la sophistication et de la variété des attaques existantes, il est beaucoup plus compliqué de traiter la robustesse d'un système que son imperceptibilité. Néanmoins, certaines conditions d'imperceptibilité doivent toujours être imposées lors de l'optimisation de la robustesse.

Il existe plusieurs façons de classer les attaques contre les systèmes de tatouage audio. Intuitivement, le tatouage audio peut être attaqué intentionnellement ou non, ce qui est déterminé par les motivations des utilisateurs correspondants. Par exemple, la

compression avec perte est susceptible d'être une attaque non intentionnelle, tandis que l'ajout de bruit peut être considéré comme une attaque intentionnelle.

Dans les sections suivantes nous montrons les classes des attaques particulièrement de traitement de signal qui sont dédiées à l'audio. Nous montrons deux classifications réalisés par des chercheurs, la première catégorise ces attaques en attaques de base et avancées, tandis que la deuxième classe les attaques par groupe en fonction du traitement réalisé sur le fichier audio tatoué. Nous montrons aussi l'impact et l'audibilité de ces attaques en fonction de divers travaux de tatouage audio.

### 4.3.1 Classification de Guang Hua et al.

Les auteurs (Guang Hua et al.) dans [68] résumant les attaques dans le tableau 4.2 et les catégorisent dans la figure 4.11. La sélection des attaques présentées dans le tableau 4.2 repose sur les considérations suivantes :

- i) Ils mettent davantage l'accent sur l'impact du traitement du signal sur le contenu audio, qui est plus susceptible de se produire du côté des attaquants.
- ii) Les attaques considérées sont une combinaison d'effets musicaux et d'attaques algorithmiques, en considérant le point de vue de Steinebach et al. dans [162].
- iii) Les attaquants sont aussi contraints de manière à ce que les attaques ne doivent pas altérer le contenu original des données audio.

Tableau 4.2 - Résumé et abréviations des attaques existantes

<b>Attaque</b>	<b>Abbrev.</b>
Boucle fermée	CLP (Close-Loop)
Requantification	RQZ (Requantization)
Filtrage	FTR (Filtering)
Mise à l'échelle de l'amplitude	ASC (Amplitude Scaling)
Compression avec perte (MP3/AAC)	LCP (Lossy Compression)
Ajout de bruit blanc Gaussien	WGN (white Gaussian noise)
Ajout d'échos	ECH (Echoes Adding)
Recadrage	CRP (Cropping)
Gigue (Jittering)	JTR (Jittering)
Décalage de temps	TSH (Time Shifting)
Mise à l'échelle temporelle invariante de la hauteur du son	TSC (Pitch-Invariant Time Scaling)
Mise à l'échelle de la hauteur de son invariante dans le temps	PSC (Time-Invariant Pitch Scaling)
Mise à l'échelle de la vitesse	SPS (Speed Scaling)
Attaque de masque	MSK (Mask Attack)
Remplacement	RPM (Replacement)

Dans la figure 4.10, les auteurs classent les attaques sélectionnées en attaques de base et avancées [68]. Les attaques de base sont généralement plus faciles à traiter que les attaques avancées.



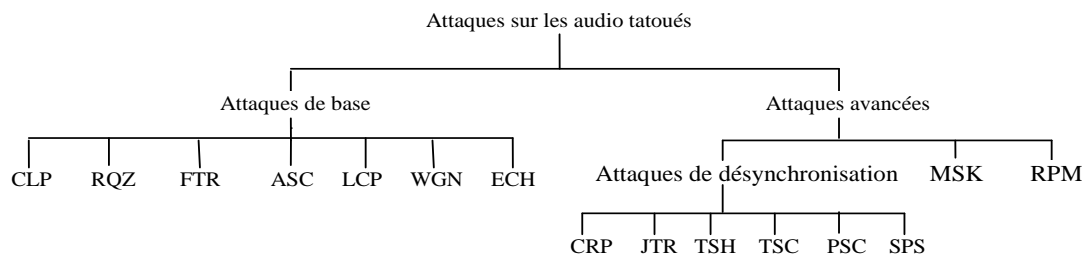


Fig. 4.10 - Catégorisation des attaques existantes du tatouage audio [68]

#### 4.3.1.1 Les attaques de base

L'attaque CLP est un environnement en boucle fermée où le signal tatoué n'est en fait pas attaqué. Cette attaque est prise en compte car elle est liée aux méthodes de non-rejet de l'interférence de l'hôte [163], dans lesquelles, pendant l'extraction de la marque, le signal hôte joue le rôle d'interférence. Les méthodes basées sur l'ajout de l'écho, l'étalement du spectre (SS) et le patchwork ne rejettent pas l'interférence de l'hôte, tandis que la méthode modulation de l'indice de quantification (QIM) rejette les interférences de l'hôte. L'attaque RQZ consiste à appliquer un niveau de quantification différent au signal tatoué, par exemple de 16 bits à 8 ou 24 bits. Les attaques FTR comprennent les attaques par filtrage passe-bas, passe-bande et passe-haut. Toutefois, l'attaque FTR ne doit pas supprimer l'information originale à fournir dans le signal tatoué. Les effets musicaux introduits par un égaliseur peuvent également être considérés comme une attaque FTR. L'attaque ASC met à l'échelle l'amplitude à un niveau différent. Les attaques LCP peuvent être des compressions MP3 ou AAC avec des débits différents mais raisonnables (par exemple, les débits inférieurs à 128 kbps peuvent être exclus car la technologie numérique actuelle d'aujourd'hui peut facilement supporter une plus grande fidélité). L'attaque WGN considère des niveaux raisonnables de bruit injecté et, en général, on limite le rapport signal/bruit résultant à plus de 20 dB. L'attaque ECH intègre des copies retardées et atténuées dans le signal tatoué.

#### 4.3.1.2 Les attaques avancées

- **Les attaques de désynchronisation :** Les attaques de base mentionnées ci-dessus ont été prises en compte dans la plupart des travaux existants, tandis que peu d'efforts ont été consacrés à une étude complète des attaques avancées. Bien que de nombreux travaux soient consacrés à l'étude des attaques de désynchronisation, les définitions de ces attaques ont été très différentes parmi les travaux existants. Une définition rigoureuse des attaques de désynchronisation devrait caractériser toutes les attaques qui introduisent des désalignements des positions de la marque entre les copies attaquées et les copies originales, ou, au moins, elle devrait inclure celles qui sont résumées dans la figure 4.10.

L'attaque CRP fait référence à un recadrage aléatoire. L'attaque JTR supprime des échantillons dans chacune des trames non chevauchantes d'une durée fixe. L'attaque TSH applique un décalage temporel aux données audio. L'attaque TSC met à l'échelle le temps du signal tout en préservant l'hauteur. Elle est également connue sous le nom d'attaque

par modification de l'échelle de temps (TSM). L'attaque PSC, quant à elle, met à l'échelle l'hauteur du son tout en préservant le temps du signal. Enfin, l'attaque SPS modifie le temps et l'hauteur du son en même temps qui est appelée aussi attaque de ré-échantillonnage. Bien que toutes ces attaques de désynchronisation puissent être facilement mises en œuvre à l'aide de lecteurs audio omniprésents (par exemple, Foobar2000, Audacity et Adobe Audition, etc.), elles ont été moins étudiées dans les travaux existants. Les attaquants peuvent facilement utiliser une combinaison de ces attaques pour contrer les systèmes de tatouage.

- **L'attaque de masque :** L'attaque MSK est réalisée en tirant parti des courbes de masquage utilisées pour contrôler l'imperceptibilité pendant la phase d'insertion de la marque. Le raisonnement est le suivant : si le modèle psycho-acoustique est mis en œuvre efficacement, la courbe de masquage du signal tatoué sera une bonne approximation de la courbe de masquage du signal hôte original. De cette façon, les marques peuvent être estimées efficacement sous la courbe de masquage du signal tatoué, et donc supprimées du signal tatoué avec un paramètre permettant d'équilibrer la distorsion de l'attaque et la force de suppression. En fait, cette attaque partage une idée similaire à celle des compressions avec perte, parce qu'elles utilisent toutes deux le modèle psycho-acoustique et suppriment les signaux dans les régions perceptivement insensibles. L'idée originale de l'attaque MSK proposée dans [164] repose sur le modèle de signal d'étalement de spectre (SS) et les estimations statistiques. Néanmoins, la même idée pourrait être appliquée à d'autres modèles d'insertion.
- **L'attaque de remplacement :** Pour les formats audio, les segments vocaux qui correspondent au même mot ou à la même voyelle d'un seul locuteur, ou les segments musicaux dans lesquels le même instrument joue le même ton, etc. peuvent tous être très similaires. Par conséquent, le remplacement de segments audio par leurs homologues similaires sur le plan perceptif dans le même signal audio entraînerait une distorsion d'attaque marginale, mais des marques totalement désordonnées. L'attaque RPM peut être appliquée à n'importe quelle méthode de tatouage, puisque la seule exigence pour effectuer cette attaque est le signal tatoué. En utilisant la distance euclidienne au carré comme mesure de similarité, les étapes de l'attaque RPM sont résumées comme suit
  1. Partitionner  $S_w(n)$  en blocs se chevauchants.
  2. Pour chaque bloc, trouver un ensemble de blocs les plus similaires.
  3. Combiner linéairement ces blocs pour optimiser la similarité.
  4. Remplacer le bloc résultant par le bloc original.

Les procédures ci-dessus reflètent l'idée principale de l'attaque de remplacement dans les travaux de [165], [166]. Notez que l'efficacité de l'attaque RPM dépend également de la possibilité de modification sur les étapes de l'attaque. Par exemple, le problème des moindres carrés associé à l'étape 3 peut être remplacé par une permutation aléatoire des blocs similaires découverts [167], ce qui est appelé attaque par correspondance aveugle. Il a été vérifié dans les travaux [165]–[167] que l'attaque RPM est très efficace contre les méthodes d'étalement de spectre (SS) et modulation de l'indice de quantification (QIM).

### 4.3.1.3 Effets des attaques sur la qualité de l'audio tatoué

Parmi les étapes de traitement dans les systèmes de tatouage audio, comme le montre la figure 3.11, la dégradation de la qualité audio provient en fait de deux phases, à savoir l'insertion de la marque et les attaques, respectivement [68]. La dégradation de la qualité audio au cours de la phase d'insertion de la marque est

étroitement liée à la propriété d'imperceptibilité. Les attaques de base et les attaques avancées entraînent également une perte de qualité audio. Les attaquants sont confrontés à un compromis entre la suppression réussie de la marque et la dégradation de la qualité audio. Par conséquent, un bon schéma de tatouage doit assurer que les marques insérées causent un minimum d'effets perceptibles et qu'un attaquant n'est pas en mesure d'enlever les marques sans graves dégradations. Cependant, dans certains cas extrêmes, l'attaquant considère qu'une attaque est réussie même si la qualité du son est gravement compromise.

Les attaques de base sont plus susceptibles de provoquer une perte de qualité perceptible si la force des attaques est hors de contrôle. Mais si les attaques sont soigneusement limitées, elles peuvent ne pas réussir. En revanche, les attaques avancées sont plus susceptibles de réussir sans dégradation notable de la qualité [68]. Par exemple, les attaques de désynchronisation, en particulier la TSC, la PSC et la SPS, peuvent réussir à supprimer les marques même si le rapport d'échelle est inférieur à 5%. Cela est dû en grande partie au fait que ces attaques sont des opérations hautement non linéaires, le signal tatoué après l'attaque TSC ou PSC pourrait être considéré comme étant traité par un système non linéaire. Il est donc très difficile de récupérer un signal passé par un processus non linéaire avec des opérations linéaires. Enfin, les attaques MSK et RPM pourraient être plus efficaces puisque la dégradation de la qualité est prise en compte lors de ces attaques, avec ces attaques les marques peuvent être supprimées sans perte de qualité notable

#### **4.3.2 Classification de Steinebach et al.**

Une contribution importante aux évaluations d'un système de tatouage audio contre les attaques a été faite par Steinebach et al. dans [162], dans lequel le système de référence *Stirmark* est proposé. Dans ce travail, les auteurs différencient les attaques en effets musicaux et en attaques algorithmiques. En plus, les attaques potentielles pendant la transmission et la lecture de données audio stockées sur des CD sont aussi discutées. Le système de référence *Stirmark* a été utilisé pour évaluer la robustesse des systèmes de tatouage audio dans plusieurs travaux de recherche.

N'importe quelle manipulation d'un fichier audio constitue potentiellement une attaque à l'encontre de la marque dissimulée. En fonction de la manière dont le son est utilisé certaines attaques sont plus probables que d'autres. Les auteurs Steinebach et al. dans [162] ont proposé d'utiliser le type de post-production en studio comme base pour ces attaques. Par exemple, la préparation d'un enregistrement sonore pour une retransmission radiophonique inclut souvent la normalisation et la compression de l'enregistrement, afin d'obtenir un niveau de volume sonore compatible avec la transmission, une égalisation pour optimiser la qualité perçue, un débruiteur ou un déhissier et différents filtres permettant de réduire les parties non désirées de l'information audio et retirer les fréquences non transmissibles (voir figure 4.11).

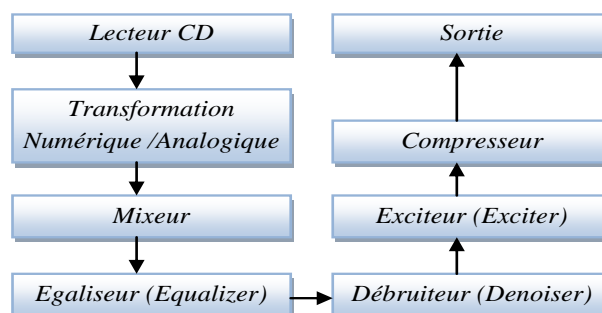


Fig. 4.11 - Transmission radio d'un signal CD joué

Un autre exemple est celui de l'Internet : si une entreprise souhaite insérer des marques pour protéger les droits d'auteur, la marque doit être robuste contre toutes les opérations habituellement effectuées sur le matériel audio. Dans ce cas, l'attaque principale sera la compression avec perte comme le MP3, parfois à des taux de compression élevés.

Pour rendre possible l'évaluation de méthodes de tatouage audio, Steinebach et al. dans [162] ont constitué des groupes d'attaques. Comme les attaques d'un même groupe reposent sur des principes identiques, il est probable qu'un schéma qui résiste à une attaque donnée dans un groupe résiste également à toutes les autres appartenant à ce même groupe (et inversement si l'algorithme y est vulnérable). Accroître la résistance à une attaque augmente donc la robustesse de la méthode à toutes les attaques du groupe considéré.

Sur la base des modèles d'attaque, les auteurs (Steinebach et al.), dans [162] ont identifié les groupes d'attaques suivants [162]:

1. **dynamique** : porte sur le profil d'amplitude d'un fichier audio. Son augmentation ou sa diminution constitue en soi une attaque. Un limiteur, un expanseur ou un compresseur sont des systèmes plus complexes car ils reposent sur des changements non linéaires dépendant du matériel audio.
2. **filtrage** : il coupe ou amplifie certaines parties du spectre. Les plus classiques sont les filtres passe-bas et passe-haut, mais les equalisers sont aussi assimilables à des filtres.
3. **ambiance** : ces effets simulent la présence d'un endroit (stade, studio, salle de concert, ou autres). Les effets les plus classiques sont la réverbération (reverb) et le retard (delay), leur paramétrage permettant différentes simulations.
4. **conversion** : le matériel audio est souvent soumis à des changements de format. Par exemple, les sons mono sont dupliqués pour donner du stéréo, la fréquence d'échantillonnage peut passer de 32 kHz à 44 kHz (voire 96 kHz), les conversions numériques/analogiques et analogiques/numériques. Toutes ces conversions impliquent du bruit et des effets sur le signal.
5. **compression avec pertes** : ces algorithmes s'appuient sur un modèle psycho-acoustique précis et permettent ainsi de réduire la taille des données d'un facteur de 10 ou mieux. Ils s'appuient sur une destruction des informations imperceptibles par un auditeur.
6. **bruit** : la plupart des attaques présentées précédemment introduisent du bruit dans le signal. Les composants matériels dans une chaîne audio injectent eux aussi du bruit dans le signal. Une attaque consiste alors à dégrader le signal en y ajoutant volontairement du bruit.

7. **modulation** : les effets de modulation, comme le vibrato, le chorus, la modulation d'amplitude ou le flanging sont rarement accessibles dans des conditions normales d'utilisation du fichier audio. Cependant, comme la plupart de logiciels les incluent, ils peuvent être utilisés comme attaques.
8. **time stretch et pitch shift** : ces opérations changent la durée d'un événement audio, sans en modifier l'hauteur, ou bien changent l'hauteur en laissant la durée intacte. Elles permettent d'obtenir un réglage précis ou de faire rentrer un signal donné dans une fenêtre temporelle.
9. **permutations d'échantillons** : Ce groupe comprend des algorithmes qui ne sont pas utilisés pour la manipulation audio dans les environnements habituels. Entre autres, il s'agit de permuter des échantillons, ou d'en abandonner.

Les deux derniers groupes appartiennent aux attaques de désynchronisation, qui provoquent un décalage entre l'encodeur et le décodeur. Par conséquent, il est difficile de récupérer une marque souffrante de telles attaques dangereuses.

Sur la base des éléments de test StirMark for Audio et SDMI, et STEP2000 deux tests de robustesse sont mis en place, à savoir un test de robustesse de base et un test de robustesse avancé [20]. Ces tests doivent être utilisés par exemple pour évaluer la robustesse d'un système de tatouage audio de droits d'auteur car ce type de système doit être robuste contre autant d'attaques que possible et pour garantir son application à différentes situations.

Rappelons que dans les tests de robustesse, les signaux tatoués attaqués ne doivent pas être dégradés au-delà des niveaux tolérés. Sur la base de ce principe, les paramètres d'attaque énumérés ci-dessous sont déterminés en conséquence.

#### 4.3.2.1 Test de Robustesse de Base

Le test de robustesse de base intègre un grand nombre d'attaques typiques sur les techniques de tatouage audio. Il est appliqué au signal audio tatoué  $S_w$  pour vérifier sa capacité à résister à différentes attaques, comme indiqué dans le tableau 3.3 qui montre les paramètres utilisés, l'expression, et la mise en œuvre de chaque attaque. Il convient de mentionner que le logiciel "Adobe Audition v3.0"<sup>3</sup> peut être utilisé pour mettre en œuvre certaines de ces attaques, autres attaques peuvent être réalisées sous MATLAB. Le test comprend les attaques suivantes [20] :

*Ajout de bruit (Noise addition)*: Ajouter du bruit blanc gaussien au signal audio tatoué pour atteindre le SNR spécifié. Habituellement, le SNR est ciblé à une valeur comprise entre 20 et 40 dB.

*Rééchantillonnage (Resampling)* : Le signal audio tatoué, dont le taux d'échantillonnage initial était de 44,1 kHz, est sous-échantillonné à 22,05 ou 11,025 kHz, puis rééchantillonné à 44,1 kHz.

---

<sup>3</sup> Adobe Audition est un puissant enregistreur, éditeur et mixeur audio numérique pour Windows. Il peut effectuer de nombreuses opérations, telles que le rééchantillonnage, la requantification, la mise à l'échelle de l'amplitude, la réverbération, la compression MPEG, l'étirement du temps et le décalage de la hauteur du son, sur divers formats de fichiers audio, .au, .voc, .vox, .wav, et ainsi de suite.

*Requantification* (Requantization) : Le signal audio tatoué qui avait à l'origine 16 bits/échantillon est requantifié à 8 bits/échantillon, puis requantifié à nouveau à 16 bits/échantillon.

*Mise à l'échelle de l'amplitude* (Amplitude scaling): L'amplitude du signal audio tatoué est remise à l'échelle par +- 10% ou +-20 %. Un taux de mise à l'échelle positif et négatif dénote que l'amplitude est amplifiée et atténuée, respectivement.

*Filtrage passe-bas* (Lowpass filtering) : Un filtre passe-bas avec une fréquence de coupure de 4, 6 ou 8 kHz est appliqué au signal audio tatoué.

*Ajout d'écho* (Echo addition) : Un écho avec un retard de 100 ou 200 ms et une décroissance de 20 ou 30 % est ajouté au signal audio tatoué.

*Réverbération* (Reverberation) : La réverbération dans une grande salle vide avec un temps de réverbération<sup>4</sup> de 1s est exercée sur le signal audio tatoué.

*Compression MP3* (MP3 compression) : Le signal audio tatoué initialement au format *.wav* est compressé à un débit binaire de 48, 64, 96 ou 128 kbps (kilobits par seconde) par un encodeur MP3. Ensuite, le fichier *.mp3* est décompressé au format *.wav* par le décodeur MP3. Il convient de noter que le processus de compression/décompression entraîne non seulement une modification de l'amplitude mais aussi un déplacement entre les signaux audio tatoués et attaqués. Une certaine quantité d'échantillons quasi nuls est ajoutée au début<sup>5</sup> et à la fin du signal attaqué [169], en raison de l'organisation interne des données dans les fichiers MP3. Par conséquent, la compression MP3 se présente sous deux formes, à savoir la *Compression 1* et la *Compression 2*. Dans la compression 1, nous supprimons ces échantillons supplémentaires afin de nous concentrer sur l'effet de la modification de l'amplitude par la compression MP3. La Compression 2 est en fait une attaque combinée qui associe la compression des données et l'insertion de zéros.

*Conversion numérique-analogique/analogique-numérique* (DA/AD conversion) : Le signal audio tatoué est lu par le lecteur audio d'un ordinateur. Ensuite, le signal de lecture est enregistré en connectant le casque phone au connecteur de la carte son de l'ordinateur.

*Recouvrement d'échantillons aléatoires* (Random samples cropping) : Un certain nombre d'intervalles de 25 ms sont coupés à des positions choisies au hasard au début, au milieu et à la fin du signal audio tatoué.

*Jittering* : Le jittering est une forme de recouplement d'échantillons aléatoires effectuée de manière régulière. Pour notre signal audio tatoué, 0,1-0,2ms sur 20ms est coupé.

---

<sup>4</sup> Le temps de réverbération d'une pièce est le temps que met le son à décroître d'un certain niveau,  $\alpha$  dB une fois que la source sonore s'est arrêtée.  $T_{60}$  est lorsque  $\alpha = 60$  dB.

<sup>5</sup> Sur la base d'opérations approfondies avec Adobe Audition v3.0, on constate qu'une quantité de 1201 échantillons est ajouté au début d'un fichier audio.

*Insertion de zéros* : Un certain nombre d'intervalles silencieux de 25 ms sont insérés de façon aléatoire dans des positions choisies au hasard au début, au milieu et à la fin du signal audio tatoué.

*Modification de l'échelle de temps en fonction de la hauteur* (Pitch-invariant time-scale modification -PITSM) : L'échelle temporelle du signal audio tatoué est étirée de  $\pm 4\%$  à  $\pm 10\%$ , tandis que la hauteur du son est préservée. Un PITSM positif se traduit par une durée plus longue et un tempo plus lent, tandis que le PITSM négatif se traduit par une durée plus courte avec un tempo plus rapide.

*Modification de l'échelle d'hauteur du tempo* (Tempo-preserved pitch-scale modification-TPPSM) : L'échelle d'hauteur du signal audio tatoué est décalée de  $\pm 4\%$  à  $\pm 10\%$ , tandis que le tempo audio est préservé. Un TPPSM positif donne une hauteur de son plus élevée (*higher pitch*), tandis qu'un TPPSM négatif donne une hauteur de son plus faible (*lower pitch*).

Les cinq dernières attaques appartiennent aux attaques de désynchronisation, qui provoquent un décalage entre l'encodeur et le décodeur (Tableau 4.3 [22]). Par conséquent, il est difficile de récupérer une marque souffrante de telles attaques dangereuses, en particulier PITSM et TPPSM [20].

Tableau 4.3 - Descriptions des tests de robustesse de base

<b>Élément de test</b>	<b>Paramètres (valeurs par défaut)</b>	<b>Expression</b>	<b>Implémentation</b>
Pas d'attaque	/	Pas d'attaque	/
Ajout de bruit	snr: signal-to-noise ratio (36 dB)	Bruit (snr)	Dans MATLAB
Rééchantillonnage	$f_w$ : sous échantillonnage de la fréquence (22.05 kHz)	Rééchantillonnage ( $f_w$ )	Adobe Audition v3.0: 44.1 kHz $\rightarrow$ $f_w$ $\rightarrow$ 44.1 kHz
Requantification	$Q_w$ : requantification à 8 bits	requantification ( $Q_w$ )	Adobe Audition v3.0: 16 bits $\rightarrow$ $Q_w$ $\rightarrow$ 16 bits
Mise à l'échelle de l'amplitude	$A_s$ : taux de mise à l'échelle (10 %)	Amplitude ( $\mp A_s$ )	Adobe Audition v3.0
Filtrage passe-bas	$f_{cutoff}$ : fréquence de coupure (8 kHz)	Lp filtering ( $f_{cutoff}$ )	Dans MATLAB
Conversion Num. Anal/Anal.Num	R: mode d'enregistrement (line-in jack)	DA/AD (R)	Adobe Audition v3.0: Jouer et enregistrer
Ajout d'écho	$A_m$ : Atténuation d'amplitude normalisée (0.3) $t_d$ : temps de retard (200 ms)	Echo ( $A_m, t_d$ )	Dans MATLAB
Réverbération	$t_{reverb}$ : Temps de réverbération (1 s)	Reverb ( $t_{reverb}$ )	Adobe Audition v3.0
Compression MP3	m: taux binaire de compression (96 kbps)	Compression 1(m) Compression 2(m)	Adobe Audition v3.0: .wav $\rightarrow$ .mp3 $\rightarrow$ .wav
Recouvrement d'échantillons aléatoires	$n_c$ : nbr. of croppings (8) $t_c$ : cropped interval (25 ms)	Cropping ( $n_c \times t_c$ )	Dans MATLAB
Insertion de zéros	$n_z$ : nbr. d'insertions (8) $t_z$ : interval d'insertion (25 ms)	Inserting ( $n_z \times t_z$ )	Dans MATLAB
Jittering	$t_j$ : couper clips (0.1 ms) $t_f$ : intervalle de coupure (20 ms)	Jittering ( $t_j / t_f$ )	Dans MATLAB
PITSM	PTSM: percentage of time stretching ( $\mp 4\%$ )	PITSM ( $P_{TSM}$ )	Adobe Audition v3.0
TPPSM	PPSM: percentage of pitch shifting ( $\mp 4\%$ )	TPPSM ( $P_{PSM}$ )	Adobe Audition v3.0

#### 4.3.2.2 Test de robustesse avancé

Le test de robustesse avancé implique des attaques plus strictes que le test de robustesse de base et est spécialement conçu pour évaluer rigoureusement les algorithmes de tatouage audio proposé. Il se compose de trois parties : un test avec StirMark pour Audio, un test de collusion, et un test de tatouage multiple [19].

*Collusion*: Nous insérons  $n$  marques différentes  $W_0^{(1)}, W_0^{(2)}, \dots, W_0^{(n)}$  dans un signal hôte  $S_0$ , et nous obtenons  $n$  signaux tatoués  $S_w^{(1)}, S_w^{(2)}, \dots, S_w^{(n)}$  respectivement. Ces signaux tatoués sont ensuite combinés pour créer  $n$  signaux tatoués moyens  $S_w^{\overline{(i)}}$  ( $1 \leq i \leq n$ ) comme suit :

$$\begin{cases} S_w^{(j)} = \text{Insertion} (S_0, W_0^{(j)}), & (1 \leq j \leq n) \\ S_w^{\overline{(i)}} = \frac{1}{i} [S_w^{(1)} + S_w^{(2)} + \dots + S_w^{(i)}], & (1 \leq i \leq n) \end{cases} \quad (4.2)$$

Lors de la détection,  $i$  marques  $W_e^{(i, \cdot)}$  sont détectées du signal tatoué moyen  $S_w^{\overline{(i)}}$  individuellement :

$$W_e^{(i,j)} = \text{Detection} (S_w^{\overline{(i)}}), \quad (1 \leq i \leq n) \text{ et } (1 \leq j \leq n) \quad (4.3)$$

Où  $n = 2 \sim 4$  dans le test de robustesse.

*Tatouage multiple* :  $n$  marques différentes  $W_0^{(1)}, W_0^{(2)}, \dots, W_0^{(n)}$  seront insérées dans un signal hôte  $S_0$  de la façon suivante :

$$\begin{cases} S_w^{(1)} = \text{Insertion} (S_0, W_0^{(1)}), \\ S_w^{(i)} = \text{Insertion} (S_w^{(i-1)}, W_0^{(i)}), & (2 \leq i \leq n) \end{cases} \quad (4.4)$$

Lors de la détection,  $i$  marques  $W_e^{(i, \cdot)}$  sont détectées du signal tatoué moyen  $S_w^{(i)}$  individuellement :

$$W_e^{(i,j)} = \text{Detection} (S_w^{(i)}), \quad (1 \leq i \leq n) \text{ et } (1 \leq j \leq i) \quad (4.5)$$

Où  $n = 2 \sim 4$  dans le test de robustesse.

Dans le cas du tatouage multiple, il faut noter que la technique pour chaque tatouage peut être identique ou différente. Par exemple, la première marque est intégrée à l'aide de l'algorithme utilisé dans le tatouage, mais la seconde marque peut être intégrée à l'aide d'un tatouage de dissimulation d'écho par exemple. Dans ce cas, la méthode de détection correspondante sera utilisée pour détecter chaque marque.



### 4.3.3 Outils pour l'évaluation des schémas de tatouage audio

Les performances d'une méthode de tatouage doivent être évaluées et rendues publiques afin que les utilisateurs sachent quelle approche à employer. Plusieurs solutions sont envisageables pour la mise en œuvre de ces tests [171] :

- faire confiance à qui fournit la méthode, ainsi qu'aux résultats qu'elle affiche;
- réaliser soi-même les tests pour vérifier que la méthode répond à ses propres attentes ;
- laisser un tiers indépendant évaluer la méthode

Seule la dernière démarche permet d'obtenir un résultat objectif, à la condition impérative que la méthodologie et les outils employés soient eux-mêmes connus de tous. Ainsi, les sources du programme seront accessibles à tous, de même que divers documents détaillant le protocole expérimental. Cette approche permettra à chacun de reproduire aisément les tests réalisés sur la plate-forme publique [171].

Des outils ont été développés pour réaliser des Benchmarks des méthodes de tatouage. Chaque outil définit un protocole expérimental avec un modèle générique des tests. Au début ces outils ont été destinés à l'évaluation des schémas de tatouage des images, par la suite certains de ces outils ont été élargis pour prendre en charge les systèmes de tatouage audio et vidéo. Parmi ces outils nous pouvons citer les suivants :

#### 4.3.3.1 Certimark : Certification des techniques de tatouage

C'est un projet Européen, lancé en Mai 2001 sous la direction de C. Rollin (Société des Auteurs et Compositeurs Dramatiques SACD), et regroupant 15 partenaires, L'objectif de CERTIMARK, est de faire en sorte que les algorithmes de tatouage soient certifiés par un "certificat" international. Ce processus de certification et d'attribution sera mené par un grand représentant international des détenteurs de droits. CERTIMARK aborde ces questions avec deux directions principales d'activité [172]:

1. La conception et le développement d'une suite complète de tests de référence pour les technologies de tatouage :
  - Outil d'évaluation complet pour les images et les vidéos ;
  - Création de nouvelles attaques, élaboration de nouveaux tests pour tout type de paramètres, définition d'une nouvelle mesure qualitative pour les images.
2. Recherche sur des algorithmes de tatouage de haut de gamme :
  - Elaboration de méthodes de tatouage en profitant des compétences variées de tous les participants ;
  - Evaluations des techniques les plus efficaces pour une exploitation commerciale future.

#### 4.3.3.2 Checkmark

On trouve également le projet Checkmark. Ce programme, développé pour Matlab 6, contient de nombreux tests (compression par ondelettes, attaque par recopie par exemple) et s'appuie sur une métrique plus performante que le PSNR : la métrique

de Watson. Celle-ci prend mieux en compte la luminance et le contraste d'une image que le PSNR [173], [174].

Le projet d'analyse comparative de tatouage *Checkmark* a été lancé pour tenter de mieux évaluer les technologies de tatouage. *Checkmark 1.2* (sorti le 14 déc. 2001) ajoute les éléments suivants : support de la couleur, nouvelles applications (logo, médical et vidéo), FAQ en ligne. *Checkmark* inclut maintenant [173]:

1. De nouvelles classes d'attaques importantes : Compression par ondelettes (jpeg2000 basé sur JASPER) ; Transformations projectives, modélisation des distorsions vidéo, déformation ; attaque par copie ; attaque par suppression de modèle ; débruitage, débruitage suivi d'une remodulation perceptive ; suppression de ligne non linéaire ; attaque par collage, échantillonnage vers le bas/vers le haut, tremblement, seuillage. En outre, les classes de test sont reprogrammées à partir de *Stirmark* et comprennent : Recadrage, retournement, rotation, rotation-échelle, FMLR, accentuation, filtrage gaussien, flexion aléatoire, transformations linéaires, rapport d'aspect, changements d'échelle, suppression de lignes, JPEG.
2. Nouvelles mesures de qualité : PSNR pondéré et la mesure de Watson qui est une métrique plus performante que le PSNR. La métrique de Watson prend mieux en compte la luminance et le contraste d'une image que le PSNR.
3. Sortie dans un format XML flexible et génération de tableaux de résultats HTML.
4. Évaluation axée sur l'application, en particulier une application non géométrique permettant de tester rapidement les algorithmes qui n'incluent pas de mécanisme de synchronisation.
5. Accès facile aux attaques individuelles dans Matlab pour les tests.

Ces deux outils *Certimark* et *Checkmark* sont complémentaires, en particulier par le grand nombre de tests différents qu'ils proposent. Par ailleurs, ils offrent la possibilité de travailler dans des environnements différents, tant par le langage utilisé (Matlab ou C++) que par le support des média (images uniquement, ou bien également des sons ou vidéos) [171].

#### 4.3.3.3 Test SDMI

L'initiative "Secure Digital Music" (SDMI) vise à développer une technologie ouverte qui protège la lecture, le stockage et la distribution de la musique numérique. Pour atteindre l'objectif d'application des droits d'auteur, SDMI intègre des marques robustes (plus précisément appelés données de contrôle de tatouage, SCD) dans le fichier musique. Ensuite, Les dispositifs conformes au SDMI, équipés de détecteurs de marques [175], peuvent ensuite identifier l'état des marques pour effectuer les opérations appropriées.

Dans [176], les facteurs techniques clés pour l'évaluation ont été définis en fonction de la performance revendiquée de la technologie, notamment l'inaudibilité, la robustesse, la fiabilité, l'efficacité de l'exploitation et l'effet sur la capacité de compression du contenu. Tout d'abord, le contenu contenant du SCD doit être perçu comme étant statistiquement indifférenciable du contenu avant l'ajout de SCD. Pour la robustesse, le contenu tatoué doit résister à un certain nombre de processus de signalisation. Le contenu est censé être échantillonné à 44,1 ou 48 kHz et quantifié à 16 bits. De plus, les probabilités de faux négatifs et de faux positifs de fiabilité ne

doivent pas être inférieures à  $10^{-2}$  et  $10^{-12}$ , respectivement.. Troisièmement, on s'attend à ce que la technologie devrait connaître des défaillances limitées après une attaque réussie et devrait également fournir une méthode raisonnable pour se recouvrir de cette attaque. La technologie doit fonctionner sur un certain nombre de plates-formes afin d'estimer l'efficacité de l'opération, c'est-à-dire mesurer le temps nécessaire pour détecter ou intégrer un SDC par rapport à la durée de l'extrait. Enfin, la technologie ne doit pas interférer avec la capacité des algorithmes de compression standard à maintenir un niveau de fidélité à des débits binaires standard.

#### 4.3.3.4 STEP 2000

STEP 2000 [177] est un projet international commun d'évaluation de la technologie de tatouage audio numérique, entrepris par JASRAC1 et NRI2 en collaboration avec les associations internationales de sociétés de gestion des droits d'auteur, la CISAC et le BIEM. Il s'agit du premier travail de ce type lancé par des organismes de gestion des droits d'auteur. L'objectif de STEP 2000 est de "certifier l'aptitude des technologies de watermark numérique en vue d'en promouvoir son utilisation". Des réponses enthousiastes de la part de nombreuses entreprises ont été reçues, ce qui a permis de réaliser une évaluation approfondie et importante de ces technologies de tatouage audio.

L'évaluation des technologies de tatouage numérique soumises a porté principalement sur deux aspects, à savoir l'audibilité et la robustesse.

1. *Audibilité* : les professionnels peuvent-ils percevoir si des marques ont été intégrées dans de la musique jouée dans un studio d'enregistrement ?

Un test d'écoute subjectif, le test ABX, a été effectué pour évaluer la qualité perceptive. Tout d'abord, l'auditeur écoute un enregistrement sonore sans marque (A), un enregistrement sonore avec marque incorporée (B), et un enregistrement sonore qui est l'un des deux (X). Ensuite, l'auditeur écoute alternativement A et B deux fois pendant 40 secondes chacune, et réécoute X pendant 40 secondes. Ensuite, l'auditeur décide si X est A ou B.

2. *Robustesse* : les données tatouées peuvent-elles être extraites après divers processus d'utilisation de la musique ?

Les tests de robustesse ont été effectués dans les mêmes conditions pour toutes les technologies soumises. En général, les données tatouées ont été manipulées d'une manière ou d'une autre, par exemple traitées dans un studio de diffusion (et un environnement de radiodiffusion potentiel), traitées pour être distribuées sur l'Internet et d'autres réseaux, et traitées par des équipements grand public couramment disponibles.

#### 4.3.3.5 StirMark Benchmark

En novembre 1997, la première version de StirMark a été publiée en tant qu'outil générique permettant de tester simplement la robustesse des algorithmes de tatouage d'images. Elle introduisait des distorsions géométriques bilinéaires aléatoires pour

désynchroniser les algorithmes de tatouage. Plusieurs versions ont suivi, améliorant l'attaque originale mais introduisant également une plus longue liste de tests. En janvier 1999, un premier benchmark a été rendu possible avec la publication de StirMark 3.1. L'extension naturelle de ce travail était un service public indépendant automatisé avec des profils d'évaluation étendus pour évaluer rapidement les bibliothèques de tatouage. C'était l'objectif du projet StirMark Benchmark Service. Une première brique pour cet outil a été le nouveau moteur d'évaluation StirMark Benchmark qu'on peut le télécharger. La version précédente de StirMark a été complètement réécrite afin qu'on puisse facilement brancher la bibliothèque de tatouage pour qu'elle soit évaluée à l'aide d'un profil d'évaluation (essentiellement une liste de tests et d'images) qu'on spécifie. Lors de la réécriture, les développeurs ont également séparé les différents composants du moteur afin qu'un utilisateur puisse coder ses propres attaques. L'utilisateur peut envoyer le code à condition qu'il soit conforme à la licence StirMark Benchmark et pour qu'il soit inclut dans la distribution principale [162], [171].

Le schéma de marquage est fourni par l'utilisateur sous la forme d'une bibliothèque de fonctions (binaire). Cette bibliothèque exporte en particulier une fonction d'information qui est utilisée pour sélectionner le profil d'évaluation à utiliser. Le profil d'évaluation est composé d'une liste de tests ou d'attaques à appliquer et d'une liste d'objets multimédia requis pour le test et triés par types et catégories. Tous les résultats des tests sont téléchargés sur un serveur SQL connecté à un serveur Web.

StirMark Benchmark 4.0 est disponible gratuitement sous forme de code binaire et de code source C/C++. Il comprend StirMark for Audio qui est un outil générique de test de robustesse pour les systèmes de tatouage audio. Il fait parti de StirMark1, un Benchmark de référence pour le tatouage d'images. Un certain nombre d'attaques ainsi que leurs paramètres sont inclus dans StirMark for Audio v0.2.<sup>6</sup>

#### 4.4 EVALUATION DE L'IMPERCEPTIBILITE

Pour concevoir des schémas de tatouage audio imperceptibles, l'imperceptibilité doit être quantifiée et mesurée. Un tatouage inaudible est la première contrainte d'un tatouage audio. Le signal audio tatoué doit être identique à l'original de point de vue perceptif. Donc il faut mesurer la quantité de déformation du signal tatoué que l'auditeur percevra par rapport à la version originale du signal. Il existe deux types d'évaluation pour l'imperceptibilité : évaluation subjective et évaluation objective. L'évaluation subjective utilise des méthodes non mathématiques basées sur des tests d'écoute pour simuler le système auditif humain (HAS), ces tests sont longs et coûteux, par conséquent ils sont utiles au stade final du développement de l'algorithme du tatouage audio. Pour l'évaluation objective, il existe deux types de mesures : les mesures objectives simples basées sur le SNR et qui ne sont pas toujours suffisantes pour révéler l'impact des effets ajoutés sur la perception humaine et les mesures perceptives objectives qui peuvent imiter la fonction du système auditif humain [27] en utilisant des modèles auditifs humains bien conçus. Ces mesures peuvent être

---

<sup>6</sup> STIRMARK V3.1 EST UN PREMIER BENCHMARK POUR LE WATERMARKING D'IMAGES PUBLIÉ EN 1999. LA DERNIÈRE VERSION EST STIRMARK 4.0, DISPONIBLE À L'ADRESSE [HTTP://WWW.PETITCOLAS.NET/FABIEN/WATERMARKING/STIRMARK/](http://www.petitcolas.net/fabien/watermarking/stirmark/)

réalisées au début et au milieu de la phase du développement de l'algorithme du tatouage.

#### 4.4.1 Evaluation subjective

Comme le récepteur d'un signal audio tatoué est le HAS, le jugement subjectif de la qualité ou de la distorsion du signal audio tatoué est le moyen ultime d'évaluer la qualité perceptive. En plus, le résultat du test subjectif peut être utilisé pour calibrer les mesures objectives. Par exemple, l'analyse de corrélation entre les mesures subjectives et le rapport signal/bruit (SNR) pondéré en fréquence peut être utilisée pour déterminer les paramètres optimaux.

La mesure subjective consiste à réaliser un ensemble de tests d'écoute en utilisant un grand nombre d'auditeurs pour établir de manière fiable la transparence du tatouage. Plusieurs idées de tests sont proposées et mises en œuvre afin d'évaluer subjectivement un fichier audio tatoué. Plusieurs modes de mesures subjectives sont utilisées dans le domaine du tatouage audio, Le test 2AFC [179] est utilisé pour déterminer la courbe de masquage en fonction des réponses des auditeurs entre l'audio original et les versions tatouées avec différents niveaux d'intégration. Les auteurs (M.Unoki et al.) dans [180], ont proposé une autre méthode de mesure de l'imperceptibilité subjective appelée test post hoc avec analyse de la variance ANOVA<sup>7</sup> (Analysis of Variance) [68]. Cette section passe en revue plusieurs mesures subjectives comme le test ABX, le score d'opinion moyen (*mean opinion score*) MOS et le grade de différence subjective (*subjective difference grade*) appelé score SDG [27].

##### 4.4.1.1 Évaluation de la transparence : ABX

Pour l'audio de haute qualité, comme les CD de musique, le signal audio tatoué doit avoir nécessairement une qualité "transparente". "Transparent" signifie que l'auditeur ne peut percevoir aucune différence entre le signal audio original et le signal audio tatoué. Le test ABX peut être utilisé dans un tel contexte.

Le test d'écoute ABX est le plus employé pour l'évaluation de la qualité audio des signaux tatoués. Il a été employé avec succès pour la mesure subjective des signaux audio. Dans le test ABX, trois signaux sont présentés à l'auditeur : le signal audio original (marqué A), le signal audio tatoué (marqué B) et un signal X choisi au hasard parmi A ou B. L'auditeur doit ensuite déterminer si X est A ou B. Dans le test, on demande à l'auditeur par exemple de noter deux clips audio en utilisant un nombre parmi {0, 1, 2, 3}, où 0 correspond à « exactement le même » et 3 correspond à « complètement différent ». Les deux clips sont un clip original et un autre choisi au hasard parmi les clips originaux et tatoués. Ensuite, l'ANOVA est appliquée aux scores donnés par tous les auditeurs [68].

---

<sup>7</sup> Le test ANOVA pour déterminer si la variation d'une mesure résulte du contenu du signal de couverture ou de la présence d'un message caché. L'ANOVA est une technique générale de test d'hypothèse statistique utilisée lorsqu'on veut déterminer si un certain nombre de groupes de données sont statistiquement différents ou non.

Le taux de réponses correctes  $r$  peut être utilisé pour décider si le signal audio tatoué est de qualité "transparente". Si  $r$  est supérieur à un seuil, par exemple  $\tau$ , nous pouvons affirmer avec une grande confiance que le signal audio tatoué est de qualité transparente. Si la réponse de l'auditeur est basée sur une estimation purement aléatoire, alors  $r$  est autour de 50% et le tatouage est transparent. Un choix populaire est  $\tau = 0,75$  [181], [182]. En général, le test ABX peut être placé dans un cadre de test d'hypothèse. Le nombre d'expériences et le seuil peuvent alors être déterminés à partir du niveau significatif [183], [184].

#### 4.4.1.2 Évaluation de la qualité absolue : MOS (mean opinion score)

Dans les applications où une certaine dégradation de l'audio est acceptable, il est nécessaire d'évaluer la qualité absolue de l'audio tatoué. Le MOS fournit une mesure absolue de la dégradation perceptive [185], où seul le son tatoué est évalué. La procédure de test, y compris l'espace de la pièce, le niveau de bruit, etc., est spécifiée dans la recommandation P.800 de l'UIT-T [186]. Après avoir écouté le son testé (c'est-à-dire le son tatoué), l'auditeur choisit un niveau sur une échelle de cinq niveaux pour évaluer la qualité du son. Le critère de choix des niveaux est indiqué dans le tableau 4.4.

Tableau 4.4 - Score d'opinion moyen

<i>MOS</i>	<i>Qualité</i>	<i>Description de la dégradation</i>
5	Excellente	Imperceptible
4	Bonne	Perceptible mais non gênante
3	Moyenne	Légèrement gênante
2	Mauvaise	Gênante
1	Très mauvaise	Très gênante

Les auteurs dans [187] ont utilisée le test subjectif MOS pour évaluer l'imperceptibilité de leurs fichiers audio tatoués. Ils ont utilisé une technique pour sécuriser une communication audio basée sur les transformations en ondelettes discrète DWT et rapide de Fourier FFT de domaine fréquentiel pour générer un tatouage audio robuste pour la protection de droit d'auteur. La marque est générée en utilisant une séquence aléatoire de bruit et insérée dans les coefficients de faible amplitude de la bande haute fréquence du spectre d'amplitude FFT. Ils ont procédé comme suit :

1. Ils ont fait écouter des fichiers audio originaux de différents types tels que classique, rock, pop, flûte, parole, jazz, instrumentaux devant différentes personnes.
2. Après cela, des fichiers audio tatoués sont joués.
3. Les auditeurs ont été invités à trouver la différence basée sur la qualité audio du fichier original et du fichier tatoué
4. Après cela, les auditeurs ont été invités à donner une note de chaque échantillon sur une échelle de 1 à 5, comme indiqué dans le tableau 4.4.
5. A partir de la note standard donnée, la note moyenne pour chaque fichier audio est calculée (voir tableau 4.5).

Tableau 4.5 - Valeurs expérimentales d'une mesure MOS subjective

<i>Exemple d'échantillon audio</i>	<i>Moyenne MOS</i>
Classical	3.9
Rock	4.2
Pop	4.7
Flute	4.2
Speech	4.1
Jazz	4.8
Instrumental	4.6

#### 4.4.1.3 Évaluation des petites dépréciations : SDG (subjective difference grade)

Alors que le MOS donne la mesure absolue du signal audio tatoué, il est souvent souhaitable de mesurer les petites distorsions du signal audio tatoué. La norme ITU-R BS. 1116 fournit une procédure pour évaluer la qualité subjective lorsque l'altération est censée être faible<sup>8</sup>.

La procédure suit l'approche "double aveugle, triple stimuli avec référence cachée". Au cours de chaque tour, trois fichiers audio (triples stimuli) sont présentés à l'auditeur : un signal de référence A, et deux signaux de test B et C. Le signal original est toujours marqué comme A. L'un des signaux B et C est le signal audio original (référence cachée) et l'autre est le signal tatoué. Les deux signaux de test sont permutés avant d'être présentés à l'auditeur. Le "double aveugle" signifie que ni l'administrateur ni l'auditeur ne sait lequel des signaux B et C est le signal de référence (le signal original). Après avoir écouté les trois signaux, l'auditeur est invité à évaluer la qualité de B et C par rapport à celle du signal original de référence connu A, respectivement. Comme le montre le tableau 4.6, la notation se fait sur une échelle continue de 1,0 à 5,0 avec une précision recommandée de 0,1. Puisque l'un des signaux B et C est le signal de référence, au moins l'un des signaux B et C doit être évalué à 5.0. Au moins 20 sujets sont requis, et chaque session de notation comprend 10 à 15 essais, chaque signal d'essai ayant une durée comprise entre 10 et 20 secondes, le score du signal de référence caché SHR et le score du signal tatoué SW sont rassemblés. Après avoir rassemblé les données brutes, le SDG est calculé comme suit :  $SDG = SW - SHR$  [178] .

La dernière colonne du tableau 4.6 montre les valeurs SDG en fonction de l'échelle de dépréciation. Ainsi,  $SDG = 0$  signifie que le tatouage est imperceptible, tandis que  $SDG = -4,0$  correspond à une distorsion très gênante. Le résultat des tests de l'essai est généralement présenté en utilisant la moyenne SDG avec l'intervalle de confiance de 95% pour chaque type de pistes d'essai. Une analyse statistique supplémentaire, telle que l'ANOVA peut être effectuée pour vérifier si les moyennes des différentes pistes sont égales [27].

<sup>8</sup> La norme ITU-R BS.1116 n'est recommandée que lorsque la dégradation attendue du signal audio tatoué est faible. Pour un signal audio tatoué de qualité intermédiaire, la norme ITU-R BS. 1534 convient mieux.

Tableau 4.6 - Echelle de dégradation à cinq notes et valeur de la SDG associée

Score	Qualité	Description de la dégradation	SDG
5.0	Excellente	Imperceptible	0.0
4.9 – 4.0	Bonne	Perceptible mais non gênante	-1.0
3.9 – 3.0	Moyenne	Légèrement gênante	-2.0
2.9 – 2.0	Mauvaise	Gênante	-3.0
1.9 – 1.0	Très mauvaise	Très gênante	-4.0

Bien que le test d'écoute subjectif soit une approche efficace pour évaluer la qualité perceptive des signaux audio tatoués, il n'est pas possible d'évaluer la qualité de l'écoute, ce test est souvent long et coûteux. Par conséquent, il est souvent utilisé au stade final du développement de l'algorithme de tatouage audio. Au début et au milieu du développement de l'algorithme, des mesures non subjectives pour l'évaluation de la qualité perceptive sont utilisées.

## 4.4.2 Evaluation objective

Le but essentiel des mesures objectives est de remplacer les tests subjectifs coûteux et longs par des mesures mathématiques simples et efficaces. Ces mesures objectives accompagnent le développement et le test des schémas du tatouage en cours de leur réalisation. Différentes mesures ont été proposées dans la littérature. Ces mesures évaluent la distorsion entre le signal tatoué  $S_w(n)$  et le signal original  $S(n)$  par des métriques différentielles ou des métriques de corrélation. Dans le traitement des signaux audio, les métriques différentielles sont les plus répandues. Pour les mesures objectives, le modèle psycho-acoustique ou le modèle auditif n'est pas explicitement utilisé. Au lieu de cela, elles exploitent le concept de SNR.

### 4.4.2.1 SNR (Signal to Noise Ratio)

Le rapport signal sur bruit est un paramètre utilisé pour connaître la valeur par lequel le signal est corrompu par le bruit. Il est défini comme le rapport de la puissance du signal à la puissance de bruit.

$$SNR = \text{Puissance du signal} / \text{Puissance du bruit} \quad (4.6)$$

Lors d'un tatouage, l'imperceptibilité de la marque insérée est mesurée par le calcul du SNR entre le signal audio original et le signal audio tatoué. La puissance moyenne des signaux peut être un indicateur de la distorsion introduite par le tatouage. Soit  $S(n)$  le signal audio hôte et  $S_w(n)$  le signal audio tatoué. Le SNR défini ci-dessous peut être utilisé comme mesure objective :

$$SNR = 10 \log_{10} \frac{\sum_{n=0}^{N-1} S^2(n)}{\sum_{n=0}^N [S(n) - S_w(n)]^2} \quad (4.7)$$

Où  $N$  est la taille de l'échantillon des signaux audio. En raison de l'effet de calcul de moyenne, le SNR global pourrait être faible même s'il existe de grandes différences locales entre  $S(n)$  et  $S_w(n)$ . Il est démontré que le SNR est en mauvaise corrélation avec les scores d'évaluation subjective [29]. Par conséquent, certaines mesures de SNR modifiées, qui pondèrent différemment les différentes bandes de fréquences, sont utilisées pour mesurer plus précisément l'imperceptibilité. L'un des



exemples est le rapport signal/bruit segmentaire pondéré en fonction de la fréquence (*fwsSNR*- *frequency-weighted segmental signal-to-noise ratio*) [188], qui s'écrit comme suit

$$fwsSNR = \frac{10}{N_{Seg}} \sum_{j=0}^{N_{Seg}-1} \log_{10} \left[ \sum_{n=N \times (j-1)+1}^{M \times j} \frac{S^2(n)}{[S(n) - S_w(n)]^2} \right] \quad (4.8)$$

Où  $N_{Seg}$  est le nombre de trames non chevauchées du signal original et du signal tatoué,  $M$  la taille de la trame, et  $n$  est l'indice de la trame.

### 4.4.3 Evaluation objective perceptive

#### 4.4.3.1 Le grade de différence objective : ODG (Objective Difference Grade)

Les mesures de SNR ne prennent pas en compte les propriétés du système auditif humain. Il est nécessaire donc d'avoir une mesure objective fiable, qui prend en considération les caractéristiques du système auditif humain. Les mesures perceptives objectives sont des mesures objectives (computationnelles) qui utilisent des modèles psycho-acoustiques ou/et de cognition de niveau supérieur. Plusieurs algorithmes de mesures objectives ont été proposés comme la mesure PEAQ (*Perceptual Evaluation of Audio Quality*), normalisée dans la norme BS. 1387 [189], et la mesure plus récente PEMO-Q (*Perceptual Model Quality assessment*) [190]. L'objectif de ces algorithmes est de quantifier la distorsion introduite par le tatouage en donnant une note ODG (*Objective Difference Grade*) associé à un attribut de dégradation (voir tableau 4.7). Comme indiqué dans [191], l'ODG du PEAQ est bien corrélé avec le score SDG du test d'écoute subjective affiché dans le tableau 4.6.

Tableau 4.7 - Echelle de dégradation à cinq notes et valeur de l'ODG associée

ODG	Qualité	Description de la dégradation
0.0	Excellente	Imperceptible
-1.0	Bonne	Perceptible mais non gênante
-2.0	Moyenne	Légèrement gênante
-3.0	Mauvaise	Gênante
-4.0	Très mauvaise	Très gênante

#### 4.4.3.2 L'algorithme PEAQ

La mesure PEAQ (*Perceptual Evaluation of Audio Quality*) a été largement utilisée dans les systèmes de tatouage audio pour évaluer la qualité perceptive [161], [192]. Le schéma fonctionnel du PEAQ est illustré à la figure 4.12 [189]. Pour chaque paire de signaux audio original et tatoué, la sortie du PEAQ est une valeur ODG qui a la même spécification que l'échelle SDG. Tout d'abord, les signaux audio sont transposés dans le domaine fréquentiel à l'aide de la DFT et/ou d'un ensemble de filtres, ensuite une analyse psycho-acoustique des composantes fréquentielles est effectuée à l'aide de la méthode du seuil de masquage et de la méthode du motif d'excitation.

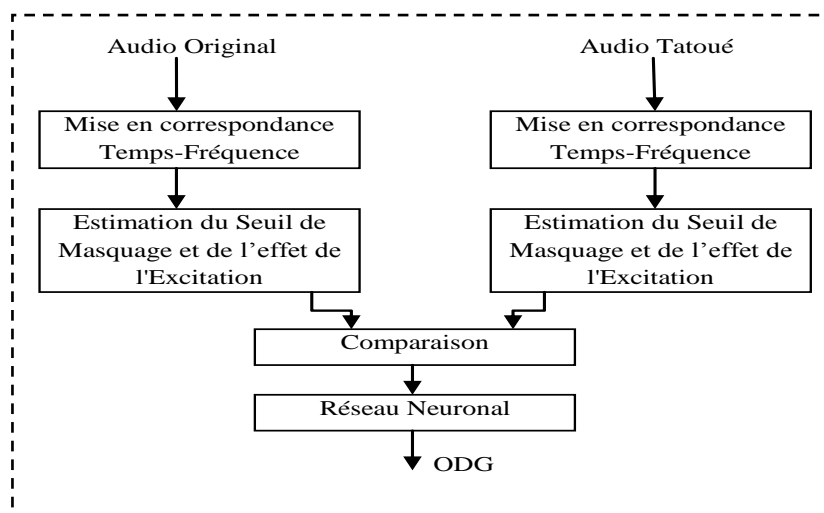


Fig. 4.12 - Schéma fonctionnel de PEAQ

Pour la première méthode, le seuil de masquage du signal audio original est analysé et comparé au spectre du signal audio tatoué. Les composantes inaudibles, c'est-à-dire celles qui se situent en dessous du seuil de masquage, sont identifiées. Dans la deuxième méthode, le schéma d'excitation de chaque signal audio sur la cochlée est estimé. Ensuite, les représentations internes sont comparées afin d'obtenir un ensemble de valeurs de sortie du modèle MOV (*model output value*) telles que l'intensité du bruit, le rapport bruit/masquage, la moyenne des blocs déformés, etc. Enfin, un réseau neuronal est utilisé pour prédire le score SDG, ou de manière équivalente la valeur ODG, à partir des *MOV*s.

Le PEAQ permet donc de classer les différences perceptives entre le signal audio original et le signal tatoué. La valeur ODG se situe dans une fourchette de  $[-4, 0]$ , comme le montre le tableau 4.7, la valeur 0 indique que les deux signaux sont perceptiblement identiques et -4 indique que les différences perceptives entre les deux sont "très gênantes". Par conséquent, plus le score ODG est proche de zéro, plus il est probable que les signaux soient perçus comme identiques.

La corrélation entre le PEAQ et le test d'écoute subjectif a été étudiée. Il a été constaté que le coefficient de corrélation entre les deux était de 0,837 et 0,851 pour la version de base et la version avancée du PEAQ respectivement. Il est certain que le PEAQ ne peut pas remplacer complètement le test d'écoute subjectif, mais c'est une mesure objective largement acceptée de la qualité audio dans l'industrie et il a été largement utilisé pour évaluer l'imperceptibilité des algorithmes de tatouage.

En résumé, nous pouvons dire que la qualité objective et perceptive peut être mesurée à l'aide des paramètres : Le rapport signal/bruit (SNR) et la note de différence objective (ODG). Selon les recommandations de l'IFPI, la marque insérée devient imperceptible lorsque le SNR est supérieur à 20 dB. Le SNR a été largement utilisé pour mesurer la qualité du signal tatoué.

#### 4.4.3.3 L'imperceptibilité pour le contrôle des marques insérées

Les méthodes présentées ci-dessus peuvent être utilisées non seulement pour évaluer la propriété d'imperceptibilité du système, mais aussi pour contrôler automatiquement la robustesse des marques insérées. Ceci est généralement réalisé en établissant un mécanisme de réglage heuristique en utilisant un coefficient  $\alpha$ , et un exemple typique est montré à la figure 4.13. Plus précisément, une valeur provisoire de  $\alpha$  est d'abord utilisée pour générer le signal tatoué. Ensuite, la procédure de mesure décide si les marques dépassent une certaine tolérance prédéfinie. S'il s'avère être audible, alors  $\alpha$  est réduit en conséquence. Sinon,  $\alpha$  est augmenté. Ce processus est répété jusqu'à ce que le niveau de distorsion soit dans la fourchette souhaitée [68].

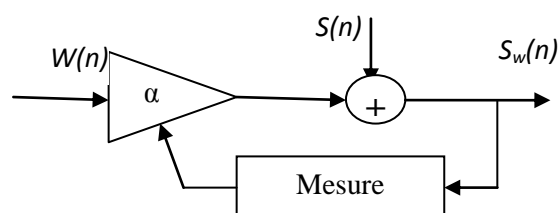


Fig. 4.13 - Schéma fonctionnel heuristique de détermination de  $\alpha$  dans un contrôle d'imperceptibilité

Une approche plus proactive pour contrôler l'imperceptibilité consiste à déterminer analytiquement la valeur de  $\alpha$  (ou à certains moments avec  $w(n)$ ), avant l'intégration de la marque, selon certaines règles d'imperceptibilité, comme l'illustre la figure 4.14. L'élément clé du module d'analyse de la figure 4.14 est le modèle psycho-acoustique du HAS, y compris la courbe de masquage et le seuil absolu d'audition ATH (*the absolute threshold of hearing*). L'imperceptibilité est donc obtenue en intégrant les marques sous le seuil de masquage ou l'ATH, de sorte qu'elles ne puissent pas être perçues par l'oreille humaine [68].

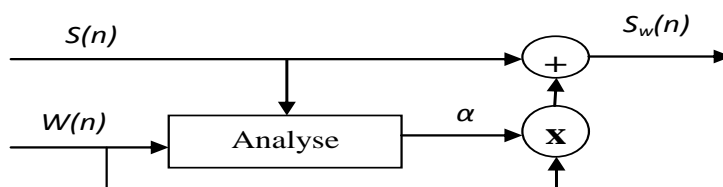


Fig. 4.14 - Schéma fonctionnel analytique de détermination de  $\alpha$  dans un contrôle d'imperceptibilité

#### 4.4.4 Les régions d'insertion des marques

Une question primordiale pour les concepteurs de systèmes de tatouage audio, bien que non mentionnée explicitement dans de nombreux travaux existants, est de savoir quels sont les meilleurs emplacements pour insérer les marques. La réponse largement acceptée à cette question est donnée dans [108]. La marque doit être insérée dans des régions perceptibles, c'est-à-dire dans des bandes de fréquences basses à moyennes, de sorte que les attaquants ne puissent supprimer les marques qu'au prix de la destruction du signal hôte. Cependant, nous avons également vu des solutions avec des marques insérées dans des bandes de haute fréquence (par exemple

des solutions basées sur le masquage d'écho [181], [182], [193]). En plus, l'utilisation d'un modèle psycho-acoustique est également facultative dans les modèles de conception existants. Par conséquent, on peut classer les régions d'insertion des marques en

- régions perceptivement significatives avec des contraintes psycho-acoustiques, par exemple [194],
- régions perceptivement significatives sans contraintes psycho-acoustiques, par exemple [127], [195], [196], et
- régions perceptivement insignifiantes, par exemple, [181], [182], [193].

Il convient de noter que l'insertion de marque dans des régions à faible fréquence est en fait très similaire à celle basée sur un modèle psycho-acoustique, puisque la courbe de masquage et l'ATH dans les régions de très basse fréquence contiennent des énergies très fortes qui garantissent que les marques sont perceptiblement limitées.

En général, les préférences concernant les emplacements d'insertion de marque seraient finalement déterminées en fonction d'applications spécifiques ayant des exigences différentes sur les propriétés d'imperceptibilité et de robustesse. La dégradation de la qualité audio après l'insertion d'une marque peut être caractérisée de manière générique du point de vue des filtres numériques ou des modèles psycho-acoustiques [68].

Malheureusement, les modèles psycho-acoustiques sont également les composants clés des compressions audio avec perte, dont les formats de compression les plus largement commercialisés sont le MP3 et l'AAC. Alors que les méthodes de tatouage audio explorent l'imperfection du HAS et insèrent les marques dans des régions perceptivement insignifiantes, les méthodes de compression audio avec perte se contentent de découvrir les données "redondantes" et "non pertinentes" dans ces régions et permettent d'obtenir des débits binaires plus faibles sans compromettre la qualité audio. En d'autres termes, si une méthode de tatouage peut identifier les régions insensibles pour l'insertion de la marque, alors les algorithmes de compression avec perte peuvent supprimer les marques intégrées. Par exemple, les marques dont l'énergie est concentrée sur les régions à haute fréquence dans les méthodes basées sur l'écho sont susceptibles d'être supprimées par le traitement commun de filtrage passe-bas avec une fréquence de coupure de 16 kHz, par exemple. Par conséquent, nous pouvons observer dans ([181], [182], [197]) que les taux de détection des marques chutent radicalement en cas d'attaques de compression avec perte à 64-96 kbps, et que des résultats satisfaisants ne peuvent être obtenus que lorsque le débit binaire de compression passe à 128 kbps et plus [68].

Sur la base de l'analyse ci-dessus, on peut tirer les conclusions suivantes :

- 1) Tant que les compressions avec perte sont l'outil prédominant pour traiter les données audio, l'imperceptibilité est généralement difficile à atteindre. Par conséquent, en intégrant des marques dans un fichier audio, la qualité du son est inévitablement compromise.
- 2) Au lieu de mettre en œuvre systématiquement un modèle psycho acoustique, un réglage heuristique basé sur [179], [189] ou [149] peut être un meilleur moyen de contrôler

l'imperceptibilité. Bien que l'accord heuristique soit plus exigeant en termes de calcul, il est réalisé par des mesures directes de la perception.

- 3) Si nous nous en tenons au contrôle systématique, il convient d'introduire certains paramètres de tolérance pour quantifier la fuite spectrale de puissance des marques par rapport à la courbe de masquage ou à l'ATH.

En plus, il existe peut-être d'autres moyens de mesurer et de contrôler efficacement la propriété d'imperceptibilité et également faciliter la sélection de la région d'insertion de la marque. Par exemple, comme un signal audio est une fonction du temps, nous pouvons utiliser efficacement cette propriété et calculer l'évolution du SNR (ou des rapports psycho-acoustiques modifiés) en fonction du temps. De cette façon, nous pouvons créer des mécanismes de post-traitement qui affinent les marques intégrées en fonction du temps, de sorte que l'imperceptibilité puisse être encore améliorée. D'une manière générale, la découverte de régions pour l'insertion de la marque dans l'audio dans le domaine spatial- fréquentiel est un problème pertinent [68].

#### **4.4.5 Niveaux d'assurance possibles pour l'évaluation de l'imperceptibilité**

Les auteurs (Raynal et al.), dans [171] propose de définir un niveau d'assurance souhaité pour chaque schéma de tatouage audio pour l'évaluation complète de l'imperceptibilité. Chaque niveau correspond à un ensemble de contraintes. Pour qu'un niveau soit validé, il faut que la méthode de tatouage les vérifie toutes. Ainsi des résultats clairs sont fournis en donnant pour chaque critère, le niveau d'assurance obtenu lors de l'évaluation. Pour l'imperceptibilité dans le tatouage, le problème est similaire à celui de l'évaluation des algorithmes de compression. Le but est de quantifier les modifications subis par le média numérique lors d'une transformation. On ne cherche pas à mesurer l'imperceptibilité de la marque mais les conséquences de son insertion dans le média audio numérique. Cette question (i.e celle de l'évaluation objective vs. subjective) est analogue à celle déjà étudiée dans le cadre du codage de compression avec pertes.

Le manque d'une mesure qualitative performante est flagrant. Le SNR ou le PSNR<sup>9</sup> sont trop restrictifs. Ces mesures ne prennent pas en compte le système visuel/auditif humain.

Un niveau d'assurance bas correspond à accepter un média audio légèrement dégradé, d'un point de vue perceptif. Les diffusions audio via un réseau présentent ce niveau de qualité. Un niveau moyen rendrait les modifications du média imperceptibles dans des conditions d'utilisation grand public. Le niveau suivant correspond à une amélioration de l'environnement, comme un enregistrement studio pour du son. Par conséquent, l'imperceptibilité devrait être mesurée par un groupe d'observateurs ou d'écouteurs examinant attentivement le média tatoué dans des

---

<sup>9</sup> PSNR est une mesure de qualité très restrictive : elle ne tient compte d'aucune propriété du modèle visuel humain. Cela inclut les propriétés de masquage habituelles mais aussi la grande tolérance aux distorsions géométriques. En utilisant le PSNR, on exclut immédiatement les systèmes de filtrage basés sur les distorsions géométriques. Malheureusement, nous n'avons pas connaissance d'une métrique prenant en compte ces distorsions [198].

conditions très strictes [171]. Ces différents niveaux sont représentés dans le tableau 4.8 [171].

Tableau 4.8 - Niveaux d'assurance possibles pour l'imperceptibilité

Niveau d'assurance	Critères
Bas	SNR ou PSNR Légèrement perceptible mais pas trop gênant
Modéré	Mesure fondée sur le système visuel/auditif humain Imperceptibles dans des conditions normales d'utilisation, i.e. grand public
Haut	Différences imperceptibles lors d'une comparaison avec le media hôte original dans des conditions type studio
Extrême	Evaluation par un grand nombre de personnes dans des conditions strictes

Tant que nous ne disposons pas de mesure précise, il est difficile de mettre au point des niveaux d'assurance fiables. Les conditions d'utilisation permettent d'introduire une distinction dans l'évaluation.

## 4.5 EVALUATION DE LA ROBUSTESSE

### 4.5.1 Mesures mathématiques

Le test de robustesse donne la correspondance entre les marques de tatouage originales et les marques extraites. Un compromis entre robustesse, imperceptibilité et capacité doit être pris en compte dans la procédure d'évaluation. De nombreuses mesures de performance différentes sont utilisées pour effectuer les tests de corrélation pour différents types de procédures de tatouage (non aveugle, aveugle et semi-aveugle). Certaines de ces mesures de performance standard sont discutées dans les sous-sections suivantes :

#### 1. Le taux de détection DR (*Detection Rate*)

La robustesse peut être mesurée par le taux de détection DR (detection rate) qui est calculé par la formule suivante [61] :

$$DR = \left(1 - \frac{\text{Nombre de bits erronés de la marque}}{\text{Nombre de bits total de la marque}}\right) \times 100\% \quad (4.9)$$

#### 2. La précision P (*Precision*)

Ce taux donne la quantité de bits de la marque originale qui sont correctement appariés avec ceux extraits [61], comme indiqué dans la formule suivante :

$$P = \left(L - \sum_{i=1}^L |W(i) - W'(i)|\right) / L \quad (4.10)$$

Où W est la séquence binaire originale de la marque, W' est la marque binaire extraite et L est la longueur de la marque.

#### 3. La corrélation normalisée NC (*Normalized Correlation*)

Ce taux indique la corrélation entre la marque originale  $W$  et la marque *extraite*  $W'$ , comme montré dans la formule suivante [61]:

$$NC = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} W(x,y)W'(x,y)}{\sqrt{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} W(x,y)^2} \sqrt{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} W'(x,y)^2}} \quad (4.11)$$

Où  $x$  et  $y$  sont des coordonnées spatiales

#### 4. Le taux d'erreur binaire BER (Bit Error Rate)

Permet de mesurer la précision avec laquelle la marque est détectée après les attaques de post-traitement sur le signal tatoué. Il est donné par la formule suivante [61]:

$$BER = (\text{Nombre de bits erronés}) / (\text{Nombre de bits total de la marque}) \quad (4.12)$$

$$BER(W, W') = \frac{100 \times \sum_{i=1}^N \begin{cases} 1 & \text{si } W_i \neq W'_i \\ 0 & \text{si } W_i = W'_i \end{cases}}{N} \quad (4.13)$$

#### 5. Peak Signal to Noise Ratio PSNR et l'erreur quadratique moyenne MSE (mean square error)

Si la marque est une image a niveaux de gris, alors, pour déterminer la dégradation de l'image originale, le pic du rapport signal-bruit (PSNR) peut être utilisé. Le PSNR est donné par la formule suivante [61]:

$$PSNR = 10 \log_{10}(x^2 / MSE) \quad (4.14)$$

Où  $x$  est la variation maximale d'intensité dans l'image et MSE est donnée par la formule suivante :

$$MSE = \sum_{i=1}^M \sum_{j=1}^n (W_{ij} - W'_{ij})^2 / M.N \quad (4.15)$$

Où  $W$  est la marque insérée,  $W'$  est la marque extraite et  $i$  et  $j$  des valeurs d'index.

#### 6. Indice de similarité structurelle SSIM (Structural Similarity Index)

Cette métrique calcule la similarité entre le signal tatoué et le signal original. La valeur SSIM choisie est comprise entre -1 et 1, et lorsque la valeur SSIM = 1, cela implique que la similarité entre eux représente le maximum. Le SSIM est présenté dans (4.16) comme suit :

$$SSIM(x, y) = \frac{\langle 2u_x \times u_y + a_1 \rangle \langle 2o_{xy} + a_2 \rangle}{\langle u_x^2 + u_y^2 + a_1 \rangle \langle o_x^2 + o_y^2 + a_2 \rangle} \quad (4.16)$$

Où  $x$  et  $y$  sont deux fenêtres d'un signal,  $u_x$  désigne la moyenne de  $x$ ,  $u_y$  est la moyenne de  $y$ ,  $o_x^2$  est la variance de  $x$ ,  $o_y^2$  est la variance de  $y$ ,  $o_{xy}$  est la covariance de  $x$  et  $y$ ,  $a_1$  et  $a_2$  sont des variables utilisées pour stabiliser la division.

## 7. Coefficient de Corrélation

Le coefficient de corrélation (CORR) est couramment utilisé dans le traitement du signal et de l'image pour trouver la similarité entre deux signaux. L'équation du coefficient de corrélation est représentée dans (4.17) comme suit :

$$CORR = \frac{\sum_i \sum_j W_{ij} W'_{ij}}{h \times w} \quad (4.17)$$

où  $W_{ij}$  et  $W'_{ij}$  sont les valeurs dans  $(i,j)$  de la marque incorporée et extraite ainsi que mis à 1 si le bit de la marque est 1 sinon mis à -1, et  $h, w$  sont la hauteur et la largeur de la marque, respectivement.

### 4.5.2 Procédure de base pour le test de robustesse face aux attaques

La procédure de base de mesure de la robustesse d'un schéma de tatouage audio face aux attaques de traitement de signal, comprend trois étapes :

1. Pour chaque signal audio à tatouer, on insère une marque  $W$  dans ce signal, avec un niveau de robustesse voulu et selon l'application envisagée, tout en assurant l'imperceptibilité du tatouage et la capacité de la marque insérée. Après l'insertion de la marque, extraire la marque  $W'$  et calculer les mesures comme le SNR, la NC, le taux de détection et le BER.
2. La deuxième étape consiste à appliquer un ensemble d'attaques sur le signal audio tatoué. Ces attaques sont des opérations courantes de traitement de signal. Ce sont des attaques non intentionnelles. L'ensemble des attaques qu'on peut réaliser sur le signal audio tatoué sont énumérées comme suit :
  - Robustesse face aux attaques de filtrage passe-bas : Un filtre passe-bas avec des fréquences de coupure de 4 KHz, 8 KHz et 11,3 KHz sera appliqué aux signaux audio tatoués.
  - Robustesse face à l'ajout de bruit blanc - AWGN (*Additive White Gaussian Noise*) : un bruit blanc gaussien est ajouté aux signaux audio tatoués.
  - Robustesse face à la compression MP3 : La compression MPEG-1 couche 3 (MP3) est la technique de compression la plus utilisée pour les signaux audio. La compression MP3 est appliquée au format WAV des signaux audio tatoués à un débit de 64 kbps, 128 kbps et 256 kbps. Enfin, le format MP3 est reconverti en format WAV.
  - Robustesse face à la ré-quantification : Chaque signal audio tatoué à 16 bits a été ré-quantifié à nouveau à 8 bits et 32 bits, puis converti à nouveau à 16 bits.
  - Robustesse face au ré-échantillonnage : Les données audio tatouées initialement échantillonnées à 44 100 Hz sont ré-échantillonnées à 22050/3200/9600 Hz, puis restaurées en ré-échantillonnant à nouveau à 44 100 Hz.
  - Robustesse face aux attaques par modification d'amplitude : Diverses modifications d'amplitude (+3 dB d'augmentation, -3 dB de réduction, +10 dB d'augmentation, -10 dB de réduction, +5% d'augmentation d'amplitude et +10% d'augmentation d'amplitude) sont effectuées sur des signaux audio.

Chowdhury & Khan (2013) [199] ont utilisé Adobe Audition CS 6 pour appliquer ces attaques sur les fichiers audio tatoués. MATLAB a également été utilisé pour appliquer des modifications d'amplitude de 5 % et 10 %, ainsi qu'un bruit blanc gaussien additif (AWGN). StirMark pour l'audio (SMBA) peut être également utilisé pour tester la robustesse face aux attaques des



fichiers audio tatoués. StirMark dispose également d'une version pour les images numériques tatouées.

3. Enfin pour chaque signal audio tatoué attaqué, Extraire la marque  $W'$  et mesurer la robustesse du processus d'extraction et du tatouage. Des mesures comme le SNR, PSNR, NC et BER peuvent être utilisées pour l'évaluation de la robustesse du schéma de tatouage après chaque attaque. Lorsque la marque du tatouage est une image à niveau de gris, il est recommandé d'utiliser des mesures comme SNR et NC, dans le cas contraire il suffit de calculer NC et BER [61].

## 4.6 EVALUATION DE LA FIABILITE

### 4.6.1 Les fausses alarmes

Deux critères permettent de définir la fiabilité d'un système, la robustesse et le taux de fausses alarmes, dont il existe deux types [171] :

- *les faux positifs* : la détection d'une marque dans un média numérique est positive alors que celui-ci ne contient pas la marque recherchée (ou pas de marque du tout) ;
- *les faux négatifs* : la détection d'une marque dans un média numérique échoue alors qu'elle y est bien présente.

Deux situations conduisent à une erreur de type faux positif [171]:

- l'algorithme de détection/extraction découvre une marque dans un média initialement non marqué (non tatoué).
- l'algorithme de détection/extraction découvre une marque  $W'$  dans un média contenant en fait la marque  $W$ .

Le premier cas se mesure en considérant un ensemble de média puis en tentant d'y retrouver une même marque. Cette opération doit être répétée plusieurs fois avec des marques différentes [171]. Pour le second, nous tentons de détecter une marque  $W'$  dans le média tatoué avec  $W$ , en utilisant la même clé que celle employée pour dissimuler  $W$ . Comme la marque est un mot binaire, nous pouvons tester un nombre significatif de mots binaires situés à une distance donnée de la marque insérée  $W$ , en accroissant cette distance au fur et à mesure. Le résultat du test est présenté sous forme d'un vecteur où la composante  $i$  représente le taux de faux positifs découverts pour un mot se trouvant à une distance  $i$  de  $W$  [171].

### 4.6.2 Niveaux d'assurance possibles pour l'évaluation de la fiabilité et la robustesse

La fiabilité exige aussi que le média tatoué doit être robuste aux attaques bienveillantes et malveillantes. Le terme robustesse, selon l'application, peut avoir des sens différents Par exemple dans le cas de la protection des droits d'auteurs, le schéma de tatouage doit être résistant aux attaques qui invalident la marque, soit en la retirant, soit en la rendant illisible, des légères détériorations du signal qui ne modifient pas ou suppriment la marque peuvent être tolérées. En revanche, lorsqu'il s'agit de vérifier l'intégrité du média tatoué, une autre granularité est recherchée. Certaines applications nécessitent la détection immédiate de la moindre modification.

D'autres supportent les changements tant que ceux-ci ne dénaturent pas l'interprétation d'un document [171].

La robustesse peut se mesurer en donnant la probabilité de détection de la marque (ou son taux d'erreur) pour un ensemble de critères appropriés à l'application considérée. Ainsi, pour chaque transformation applicable au média, on en augmente la puissance, en fonction du niveau d'assurance souhaité. Comme dans le cas de l'imperceptibilité, des niveaux d'assurance peuvent être définis pour la robustesse et la fiabilité. Chaque niveau exige certains critères qu'il faut les considérer dans les tests à réaliser. Les auteurs (Raynal et al.), dans [171] donnent des détails sur ces niveaux :

- Le niveau *bas* doit empêcher des utilisateurs « honnêtes » d'altérer la marque dans des conditions normales d'utilisation du médium. L'invalidation de la marque nécessite peu de moyens. Pour des images, cela signifie que des opérations de compression, de redimensionnement ou de recadrage de l'image ne retirent pas la preuve de propriété.
- Le niveau *modéré* est atteint lorsque des outils plus complexes sont indispensables pour compromettre le schéma. En reprenant l'exemple des images, l'emploi d'un outil avancé de traitements d'images ne permettrait pas systématiquement d'altérer la marque.
- Le niveau *haut* exige, en plus d'outils spécifiques, une bonne connaissance et des compétences dans le domaine. Toutes les tentatives pour invalider la marque ne réussissent pas systématiquement,
- Le niveau *extrême* implique la mise en œuvre de moyens plus évolués démesurés, comme des recherches par un groupe de spécialistes, qui rendent le coût de l'attaque bien plus élevé que celui requis pour l'obtention du même média marqué.

La robustesse est démontrable lorsqu'un adversaire est incapable d'invalider la marque quelque soit les attaques et les outils dont il dispose. On peut comparer cela à certains algorithmes de cryptographie fondés sur des problèmes difficiles (logarithme discret ou factorisation par exemple).

## 4.7 EVALUATION DE LA CAPACITE

La capacité d'information dissimulable dans le média résulte souvent d'un compromis entre la robustesse et l'imperceptibilité de la marque. La capacité a une importance particulière car elle a un impact direct sur la robustesse et l'imperceptibilité. Une capacité d'insertion grande cause des distorsions dans le signal tatoué et par conséquent une faible robustesse du schéma de tatouage.

La capacité se mesure en nombre de bits pouvant être insérés par unité de temps. Soit  $S$  la durée en seconde du fichier audio tatoué et  $L_w$  la taille de la marque en bits. La capacité se calcule comme suit :

$$Cap = \frac{L_w}{S} \text{ bits/s} \quad (4.18)$$

Dans [200], les auteurs donnent une autre métrique pour calculer la capacité, c'est le nombre maximum de bits de données qui peuvent être insérés dans l'objet hôte sans créer des distorsions majeurs dans le média original qui affectent la visibilité ou l'audibilité des données. La capacité est présentée dans (4.19) comme suit :

$$Capacity = \frac{\text{(Nombre de bits cachés)}}{\text{(Nombre de bits total du signal hôte)}} \quad (4.19)$$

Cette mesure est également connue sous le nom de *charge utile* qui définit les limites des informations à dissimuler tout en garantissant que la marque reste robuste et imperceptible. La capacité de la marque est déterminée par les informations dont disposent l'attaquant, l'encodeur et le décodeur de données, les restrictions de distorsion et le modèle statistique employé dans la génération de la marque.

Pour estimer la capacité maximum théorique du système de tatouage nous pouvons calculer le rapport de la capacité à la fiabilité en fixant deux paramètres le premier est la taille maximale de la marque à insérer et le deuxième le taux erreur binaire (BER).

## 4.8 CONCLUSION

Dans ce chapitre nous avons présenté les propriétés du système auditif humain car la modélisation psycho-acoustique est importante dans le codage et le tatouage audio pour garantir que les modifications du signal original restent imperceptibles, nous avons présenté des classifications des attaques non intentionnelles « attaques courantes de traitement de signal » et intentionnelles pouvant être réalisées sur les systèmes de tatouage audio et les outils cités dans la littérature permettant d'évaluer ces systèmes face à ces attaques, et enfin nous avons présenté plusieurs aspects d'évaluation des systèmes de tatouage audio.

Pour l'évaluation de l'imperceptibilité, nous avons présenté les différents types d'évaluation subjective et objective. Nous avons parlé des tests subjectifs généralement élaborés par des tests d'écoute qui s'appuient sur les caractéristiques perceptuelles du système auditif humain, les mesures objectives qui utilisent des formules mathématiques comme le SNR et la corrélation normalisée NC, et les mesures objectives et perceptives utilisant également les propriétés du HAS. Pour l'évaluation de la robustesse et la fiabilité de la marque insérée, nous avons présenté une procédure qui consiste à prendre des mesures mathématiques de SNR, PSNR, corrélation NC et taux d'erreur binaire BER, sur le signal audio tatoué et la marque extraite, ensuite nous réalisons des attaques par des opérations de traitement de signal, après chaque attaque, nous récupérons la marque insérée, nous recalculons les mesures et nous comparons les résultats obtenus. L'évaluation de la capacité nous permet de voir le nombre de bits pouvant être insérés dans le signal audio original par unité de temps.

Les méthodes de tatouage sont de plus en plus nombreuses. Néanmoins, il est difficile de les comparer et de trouver celle adaptée aux besoins dans la mesure où les tests présentés sont très souvent différents, car les média employés et les transformations qu'ils subissent changent d'une étude à l'autre.

En ce qui concerne les mesures de performance, la plupart des techniques de tatouage pour lesquelles le type de support représente une image sont évaluées en fonction des valeurs MSE, PSNR, BER et SNR. En revanche, d'autres types de

supports utilisent la consommation d'énergie et le coût de calcul pour l'évaluation. En plus, la plupart des études examinées se concentrent sur la satisfaction d'un type d'exigences en matière de tatouage, telles que la sécurité, la robustesse ou la capacité. Par conséquent, la recherche ou le développement d'un nouvel algorithme de tatouage qui unifie ces exigences est considéré comme un défi potentiel. Un autre défi lié aux types d'attaques pouvant être pris en compte, il faut améliorer la technique de tatouage existante pour qu'elle puisse prendre en charge toutes les attaques possibles sur un même schéma du tatouage [65].

Les algorithmes de tatouage étant tous différents, des niveaux d'assurance sur les critères habituels (imperceptibilité, robustesse et capacité) ont été proposés en fonction de l'application, afin de faciliter la lisibilité des performances obtenues par les schémas de tatouage. Cependant il faut noter qu'il est difficile de mettre au point des niveaux d'assurance fiables tant que nous ne disposons pas de mesure précise, les conditions d'utilisation des systèmes de tatouage audio permettent d'introduire une distinction dans leur évaluation.

## **5. Nouvelles Approches de Tatouage Numérique pour la Protection des Fichiers Audio**

---

### **Sommaire**

---

5.1	<b>Introduction</b> .....	120
5.2	<b>Un algorithme de tatouage numérique audio basé sur la transformée de Fourier pour la protection des sons respiratoires</b> .....	120
5.3	<b>Approches spatiales et fréquentielles pour la protection des fichiers audio ..</b>	127
5.4	<b>Un schéma de tatouage basé sur la combinaison de valeur de parité pour la protection des sons de battements de cœur</b> .....	138
5.5	<b>Conclusion et perspectives</b> .....	147

---

## 5.1 INTRODUCTION

Dans ce chapitre, nous présentons nos contributions réalisées dans le cadre de cette thèse. Ce chapitre est organisé comme suit : Dans La section 1 nous décrivons notre première contribution intitulé "*Un algorithme de tatouage numérique audio basé sur la transformée de Fourier pour la protection des sons respiratoires*", nous montrons les expériences réalisées et nous discutons des résultats obtenus. Dans la section 2 nous décrivons notre deuxième contribution intitulée "*Approches spatiales et fréquentielles pour la protection des fichiers audio*" avec les expériences réalisées et les résultats obtenus. Dans la section 3 nous décrivons notre troisième contribution publiée comme un "*book chapter*" et intitulée "*Un schéma de tatouage basé sur la combinaison de valeur de parité pour la protection des sons de battements de cœur*" en détaillant les expériences réalisées et les résultats obtenus. Chaque approche est sanctionnée par une analyse des résultats et une comparaison avec les travaux connexes présentés dans la dernière section du troisième chapitre de cette thèse, intitulé "*étude de l'état de l'art sur le tatouage audio numérique*". Enfin une conclusion générale sur nos approches et des perspectives futures seront présentés en section 4.

## 5.2 UN ALGORITHME DE TATOUAGE NUMERIQUE AUDIO BASE SUR LA TRANSFORMEE DE FOURIER POUR LA PROTECTION DES SONS RESPIRATOIRES

### 5.2.1 Contexte d'étude et problématique

Les bruits respiratoires sont des indicateurs importants de la santé respiratoire d'une personne. Les sons émis lorsqu'une personne respire peuvent être utilisés pour diagnostiquer des problèmes respiratoires. En raison de l'évolution des maladies, plusieurs diagnostics restent insuffisants, d'où la nécessité de la coopération de plusieurs collègues afin de parvenir à un diagnostic correct. Mais le problème majeur reste l'échange de données sur Internet, tout en préservant leur intégrité et leur confidentialité face à l'émergence importante de pirates. Le transfert de données sensibles ne peut se faire avec de tels risques et doit donc être sécurisé. Plusieurs méthodes ont été déployées pour pallier à ce problème [201]. La cryptographie a été depuis longtemps l'outil efficace utilisé pour protéger et sécuriser les informations en transmettant des données chiffrées. Elle offre ainsi des solutions pour protéger la confidentialité des données, pour assurer leur intégrité ou pour connaître l'identité de la personne qui les envoie [202]. Mais ces protections n'agissent que pendant la transmission et la distribution. Une fois que les données sont en clair, elles ne contiennent plus aucune protection. C'est pourquoi d'autres techniques de protection ont été mises en place [203]. Il s'agit de techniques de dissimulation qui visent à offrir une protection permanente même lorsque les données sont en clair et donc utilisables ou exposées à certaines manipulations. La technique du tatouage vise à marquer des documents en insérant des données invisibles mais surtout robustes. Une des particularités de cette technique est que la marque est liée de manière intime et résistante au contenu du document [204]. Elle est souvent dédiée à la protection des droits d'auteur, lorsque la marque est robuste. L'imperceptibilité, la robustesse et la capacité de la marque sont les propriétés importantes qui caractérisent la technique du tatouage. La robustesse garantit que la marque invisible ne peut pas être détruite sans

que le support soit fortement dégradé [205]. La transparence ou l'imperceptibilité indique que, même en insérant des informations secrètes, le document tatoué (contenant les informations secrètes) n'est pas perceptivement perturbé et la capacité indique la quantité d'informations qui peut être insérée sans être audible [206]. Ces propriétés sont étroitement liées mais elles ont l'effet inverse. Si la capacité est grande, elle a un effet positif sur la robustesse mais un effet négatif sur l'imperceptibilité et vice versa. Ainsi, la technique dépend étroitement de la solution qui gère le compromis robustesse/imperceptibilité. Nous sommes confrontés à un défi en termes d'optimisation. Une grande majorité des travaux de recherche ont été consacrés à l'amélioration de ce compromis en mettant en œuvre des techniques de plus en plus avancées [207]. Afin d'améliorer la sécurité des données échangées en télémédecine, nous proposons dans ce travail un schéma de tatouage aveugle pour la protection des sons respiratoires. La marque est constituée des informations du patient et des données d'acquisition. Ainsi, lors du processus d'extraction, une comparaison entre les informations du patient et la marque indiquera si les données ont été modifiées. Cela permettra de s'assurer que l'échantillon correspond bien au patient indiqué dans le dossier.

Dans ce travail, nous proposons un schéma de substitution pour le tatouage audio basé sur la transformée de Fourier. Ce schéma comprendra deux variantes dans lesquelles une transformée est appliquée. La parité des coefficients obtenus sera ensuite combinée afin de cacher la marque dans la bande de moyenne fréquence. L'objectif est d'apporter le moins de modifications possibles afin de préserver l'imperceptibilité du processus de dissimulation. Une transformation inverse permet ensuite de générer le fichier tatoué. Dans nos expériences, plusieurs variantes de la transformée de Fourier sont utilisées, nous appliquerons donc une transformée de Fourier discrète, une transformée de Fourier fractionnaire et une transformée de Fourier de quaternion. Pour chaque transformée, les deux variantes de notre schéma de tatouage seront appliquées et les résultats obtenus nous permettront de déterminer quelle transformée est la plus appropriée pour notre processus d'insertion. Une comparaison des résultats obtenus avec des travaux connexes nous permettra également de situer et d'évaluer notre approche. Cette première partie de nos contributions est organisée comme suit, la section 1 décrit l'approche proposée et les deux variantes possibles. La section 2 présentera les expériences menées et les résultats obtenus. Enfin, une analyse et des discussions de notre travail seront proposées dans la section 3.

### 5.2.2 Schéma de tatouage proposé

L'approche proposée est une combinaison de codage matriciel et de correspondance LSB (LSBM). Le codage matriciel nous permettra de garantir l'incorporation du message secret avec un nombre minimal de modifications (figure 5.1). La correspondance LSB atténuera l'effet du nombre de pair (PoV) de la méthode de remplacement LSB (LSB). Les schémas de dissimulation de données proposés sont basés sur la différence des valeurs des coefficients. Les valeurs des coefficients sont comparées entre elles en fonction de la variante utilisée. La marque à insérer est une séquence de 0 et 1, ces données sont substituées au LSB des valeurs de coefficients selon la variante utilisée. Le processus d'insertion (figure 5.1) comporte cinq étapes :

1. Le fichier audio original est d'abord transformé en domaine de transformée de Fourier.

2. La partie réelle des composantes à moyenne fréquence est sélectionnée.
3. La parité des coefficients est modulée en fonction des bits de la marque.
4. Les parties réelles marquées sont combinées avec les parties imaginaires non marquées.
5. La transformation inverse est effectuée pour générer le fichier audio tatoué.

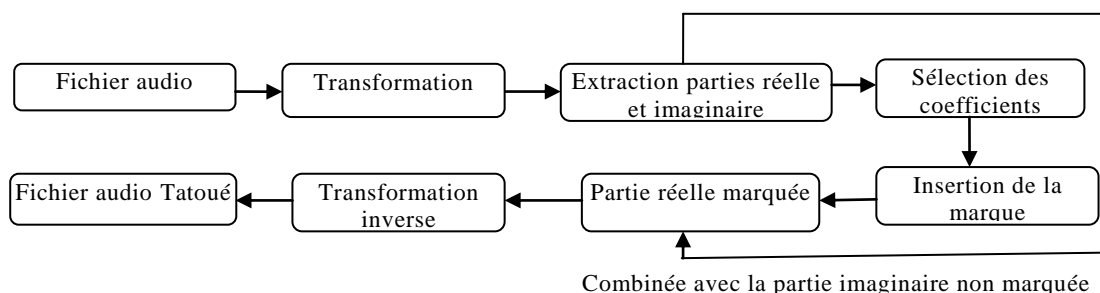


Fig. 5.1 - Le processus de tatouage

L'énergie du signal est concentrée dans les composantes de basse fréquence. L'insertion d'une marque dans ces fréquences offre une bonne robustesse, mais introduit des distorsions apparentes dans le domaine spatio-temporel. En revanche, l'insertion d'une marque dans les composantes de haute fréquence ne dégrade pas la qualité du signal, mais rend la marque vulnérable aux attaques. Par conséquent, la bande de fréquence moyenne est la plus appropriée pour l'insertion de la marque, car elle répond à l'exigence du compromis entre robustesse et invisibilité. La marque est donc insérée dans la partie réelle de ces coefficients selon les règles du tableau 5.1 ou du tableau 5.2 selon la variante utilisée. Dans la première variante, trois coefficients sont utilisés pour dissimuler deux bits de la marque, la parité de trois coefficients successifs est calculée puis comparée aux bits à dissimuler. En cas d'inégalité, la parité d'un coefficient est alors modifiée pour satisfaire l'égalité selon les règles du tableau 5.1. Considérons pour cela C1, C2 et C3 trois coefficients successifs et X, Y deux bits du message à insérer.

Tableau 5.1 - Règles de substitution pour la première variante

Condition	Action
$X = (C1 - C2) \% 2 \ \& \ Y = (C1 - C3) \% 2$	Aucun changement requis
$X \neq (C1 - C2) \% 2 \ \& \ Y = (C1 - C3) \% 2$	Changer C2
$X = (C1 - C2) \% 2 \ \& \ Y \neq (C1 - C3) \% 2$	Changer C3
$X \neq (C1 - C2) \% 2 \ \& \ Y \neq (C1 - C3) \% 2$	Changer C1

Pour la deuxième variante proposée, deux coefficients sont utilisés pour dissimuler un bit de la marque, la parité de deux coefficients successifs est calculée, et la substitution d'un bit est effectuée en appliquant l'une des règles du tableau 5.2. Considérons pour cela C1 et C2 deux coefficients successifs et X le bit du message à insérer.

Tableau 5.2 - Règles de substitution pour la deuxième variante

Condition	Action
$X = (C1 - C2) \% 2$	Aucun changement requis
$X \neq (C1 - C2) \% 2$	Changer C1

Dans la première variante, la capacité d'insertion correspondra aux 2/3 des coefficients intégrables, pour la deuxième variante, la capacité correspond à la moitié



des coefficients intégrables. Une transformation inverse sera enfin appliquée aux trois composantes afin de constituer l'objet hôte tatoué.

### 5.2.3 Résultats expérimentaux et tests

Dans cette section, les techniques de tatouage proposées sont testées sur une grande base de données de fichiers de sons respiratoires. Cette base de données est décrite par (Bruno M. Rocha et al., 2019) dans [208], ce papier décrit une base de données publique de sons respiratoires, qui a été compilée pour un concours international, le premier défi scientifique de la Conférence internationale sur l'informatique biomédicale et de santé de l'IFMBE. La base de données comprend 920 enregistrements sonores (au format .wav) acquis auprès de 126 participants et deux ensembles d'annotations. L'un des ensembles contient 6898 cycles respiratoires annotés, certains comprenant des crépitements, des sifflements ou une combinaison des deux, et d'autres ne comportant aucun bruit respiratoire fortuit. Dans l'autre ensemble, les emplacements précis de 10 775 événements de crépitations et de sifflements ont été annotés. Les cycles ont été annotés par des experts respiratoires, les enregistrements ont été recueillis à l'aide d'équipements hétérogènes et leur durée était comprise entre 10 et 90 s. Une description complète de la base de données des sons respiratoires est donnée dans l'annexe 1 de cette thèse.

Afin d'évaluer notre processus d'insertion, nous avons appliqué différentes transformées de Fourier aux fichiers audio, pour chaque résultat de transformation, les coefficients obtenus sont marqués à l'aide des deux variantes proposées. La marque utilisée dans nos expériences est une séquence de 4500 bits (représentant les informations du patient et les données d'acquisition). Dans la première expérience, une DFT est appliquée au fichier audio ce qui génère une moyenne de 13503 coefficients intégrables. La première variante nécessite trois coefficients pour insérer deux bits, ce qui permet d'insérer 9002 bits. La deuxième variante nécessite deux coefficients pour l'insertion d'un bit, et aura une capacité de 6751 bits. La figure 5.2 montre les amplitudes du signal original et du signal tatoué en modifiant les coefficients DFT.

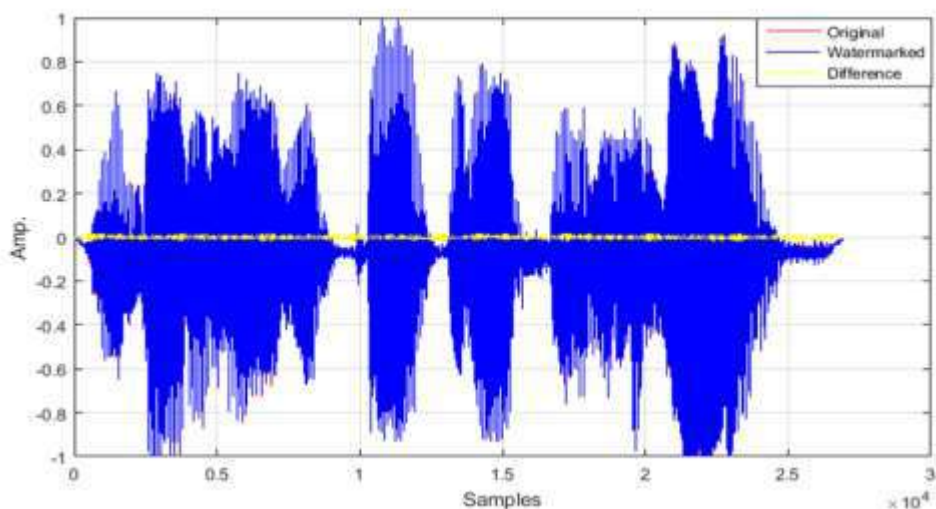


Fig. 5.2 - Représentation de l'amplitude des coefficients DFT tatoués

Pour notre deuxième expérience, les deux variantes du processus de tatouage proposé sont appliquées aux coefficients obtenus en appliquant un FrFT au fichier audio. Cela génère une moyenne de 26906 coefficients intégrables par plan, nous aurons donc une capacité de 17937 bits pour la première variante et 13453 bits pour la seconde variante. La figure 5.3 montre les amplitudes du signal original et du signal tatoué en modifiant le coefficient FrFT.

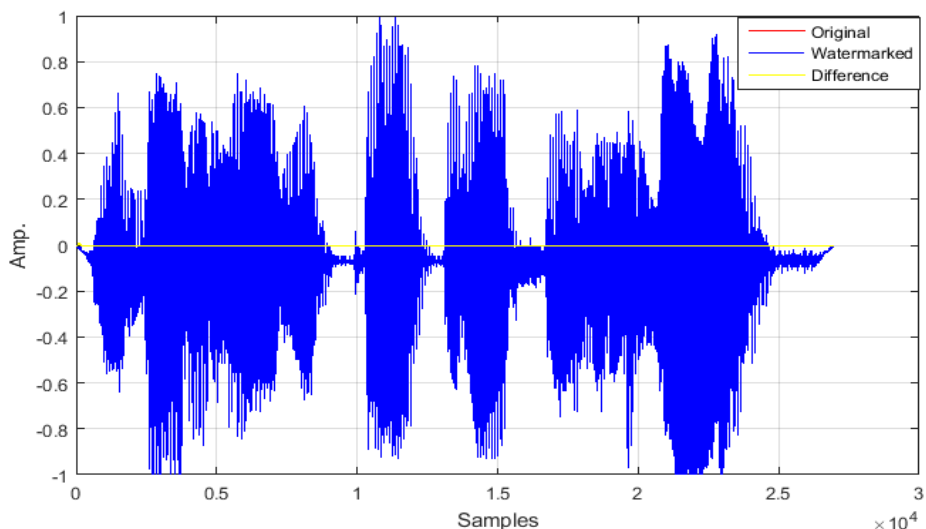


Fig. 5.3 - Représentation de l'amplitude des coefficients FrFT tatoués

Dans notre troisième expérience, l'application d'une QDFT au fichier audio permet d'obtenir 13427 coefficients intégrables. Comme la première variante nécessite trois coefficients pour intégrer deux bits, l'insertion de 8951 bits sera donc possible. La deuxième variante nécessite deux coefficients pour l'insertion d'un bit, et aura une capacité de 6713 bits. La figure 5.4 montre les amplitudes du signal original et du signal tatoué en modifiant les coefficients QDFT.

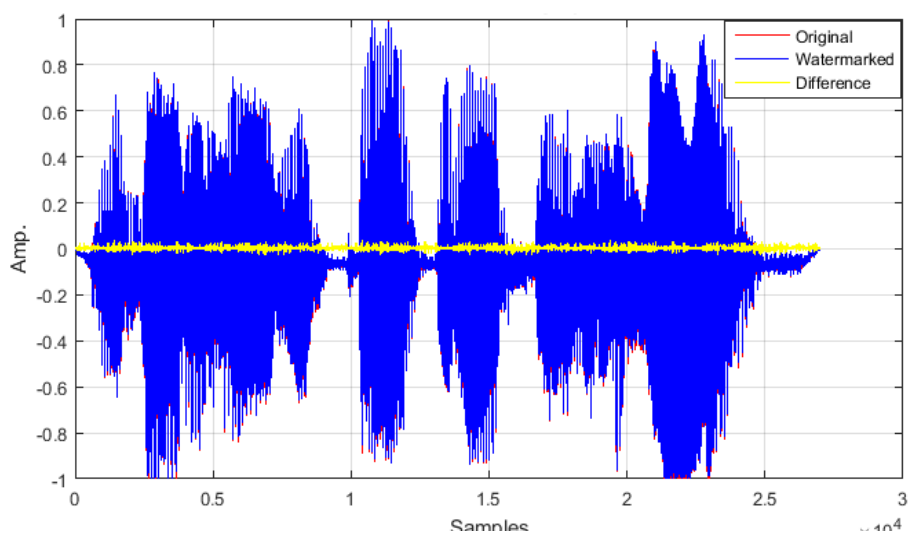


Fig. 5.4 - Représentation de l'amplitude des coefficients QDFT tatoués

### 5.2.3.1 Test d'imperceptibilité

Comme on peut le voir sur les figures 5.2, 5.3 et 5.4, les résultats obtenus ne permettent pas de détecter visuellement les distorsions générées par le processus de dissimulation. Pour cette raison, les mesures de similarité et de distorsion ont été calculées en comparant les fichiers audio originaux et les fichiers tatoués (tableau 5.3). Afin de mesurer et de tester les performances des approches proposées, la mesure SNR est utilisée. Le SNR est une métrique qui mesure la quantité de bruit ajoutée au fichier audio pendant le processus de tatouage [209]. Une valeur élevée du SNR est une caractéristique souhaitable et signifie que le niveau de distorsion est faible, le SNR est l'évaluation la plus couramment utilisée [210]. Afin d'effectuer une meilleure évaluation de chaque algorithme, nous avons également calculé l'erreur quadratique moyenne (MSE), le coefficient de corrélation (CORR), la variance (VAR) et l'écart type (STD). Dans ce tableau, DFT-V1 et DFT-V2 représentent les résultats du tatouage obtenus pour l'insertion dans les coefficients d'une Transformée de Fourier Discrète, FRFT-V1 et FRFT-V2 représentent les résultats obtenus en appliquant une Transformée de Fourier Fractionnaire. QDFT-V1 et QDFT-V2 correspondent aux résultats de tatouage obtenus en appliquant une Transformée de Fourier Discrète de Quaternion (V1 et V2 correspondent respectivement à la première et à la deuxième variante).

Tableau 5.3 - Résultats d'imperceptibilité

Variante de transformée	SNR	MSE	CORR	VAR	STD
DFT-V1	35,78	6,66000	0,99965	0,57100	0,74200
DFT-V2	37,97	9,60000	0,99980	0,78400	0,88600
FRFT-V1	35,79	6,67000	0,99990	0,57200	0,74300
FRFT-V2	37,98	9,61000	0,99990	0,78500	0,88600
QDFT-V1	37,95	9,54000	0,99950	0,78300	0,88500
QDFT-V2	35,93	6,68000	0,99970	0,57900	0,74800

Afin de comparer et de situer nos approches par rapport aux travaux récents, nous avons comparé pour chaque transformation, les résultats de distorsion obtenus avec ceux des travaux connexes présentés dans la dernière section du troisième chapitre de cette thèse, intitulé « étude de l'état de l'art sur le tatouage audio numérique ».

Tableau 5.4 - Comparaison des résultats d'imperceptibilité avec des travaux connexes

	DFT-V1	DFT-V2	FRFT-V1	FRFT-V2	QDFT-V1	QDFT-V2	Abdelwahab & al. [133]	Hu & Lee [135]	Ortiz & al. [134]	Seyed & al. [136]
SNR	35,78	37,97	35,79	37,98	37,95	35,93	35,28	28,63	35,6	39,46

Comme on peut le voir dans le tableau 5.4, les valeurs SNR obtenues par nos approches restent raisonnables par rapport aux travaux connexes. Dans la première approche, le processus de dissimulation combine trois valeurs pour insérer deux bits, en ne modifiant éventuellement qu'une seule valeur. Sur ces trois valeurs, une seule peut être modifiée, ce qui réduit la probabilité de modification contrairement aux autres approches. Cela implique moins de modification et donc génère moins de distorsion du fichier hôte, ce qui explique les bons taux de SNR obtenus et donc une imperceptibilité raisonnable. Cependant, comme notre première approche nécessite trois valeurs pour insérer deux bits, cela a un impact sur la capacité de dissimulation de l'approche. Dans le domaine spatial, cela peut être négligeable puisque toutes les valeurs des échantillons sont disponibles pour l'insertion. Mais cela devient plus

problématique dans le domaine fréquentiel où le nombre de coefficients intégrables est réduit.

### 5.2.3.2 Test de robustesse

Pour évaluer la robustesse des schémas proposés (pour chaque transformation), plusieurs attaques ont été exécutées sur les échantillons tatoués, et les résultats obtenus après l'extraction de la marque sont comparés aux travaux connexes. La corrélation normalisée (NC) est appliquée pour évaluer la robustesse du schéma de tatouage en comparant la marque originale et la marque extraite (après attaque des échantillons tatoués). Normalement, une  $NC > 0,85$  implique une similarité considérable entre la marque originale et la marque extraite.

Tableau 5.5 - Comparaison de la robustesse avec des travaux connexes

	Abdelwahab & al. [133]	Seyed & al. [136]	DFT-V1	DFT-V2	FRFT-V1	FRFT-V2	QDFT-V1	QDFT-V2
Ajout de bruit Gaussien	0,92	0.9328	0.91800	0.90543	0.90833	0.90542	0.92871	0.9109
Filtrage passe-bas	/	0.9798	0.92237	0.91735	0.96345	0.90213	0.93057	0.9446
Ré-échantillonnage	/	0.9912	0.96838	0.96082	0.98265	0.98814	0.94325	0.9873
Ré-quantification	/	0.9907	0.97752	0.97639	0.96131	0.97513	0.98521	0.9614

Comme nous pouvons le voir dans le tableau 5.5, les schémas proposés génèrent des échantillons tatoués raisonnablement robustes contre diverses attaques avec une marque de haute qualité. Habituellement, dans le domaine spatial, si les LSB sont modifiés, l'extraction de la marque devient impossible. Cependant, dans les schémas proposés, nous pouvons récupérer la marque incorporée même après la distorsion des bits LSB. Ceci s'explique par le peu de modification effectuée, dans les approches proposées la parité des éléments successifs est combinée afin d'insérer la marque, ce qui réduit la probabilité de modification. Nous pouvons donc conclure des résultats obtenus que nos approches puissent résister à plusieurs attaques malveillantes de suppression de la marque tout en garantissant une valeur SNR raisonnable.

### 5.2.4 Analyse et discussions des résultats

Dans cette contribution, deux nouvelles approches ont été proposées pour le tatouage des sons respiratoires dans le domaine fréquentiel, ce qui permet de protéger les échantillons pendant le transfert (pour préserver les informations médicales dans le fichier) et d'éviter la confusion entre les échantillons stockés puisque l'extraction de la marque garantit que l'échantillon correspond bien au patient indiqué dans l'enregistrement. Après le calcul d'une DFT, la substitution des bits à insérer a été réalisée en modifiant les LSB des coefficients quantifiés en comparant la parité des coefficients successifs. La combinaison entre les coefficients nous a permis de cacher deux bits en modifiant une seule valeur. L'insertion de la marque dans les coefficients supérieurs peut produire une distorsion importante du signal, alors que l'insertion dans les coefficients inférieurs rend la marque robuste à la compression et au filtrage. A cet effet, la marque a été insérée en utilisant les coefficients de la bande de fréquence moyenne (composantes temporelles et fréquentielles mixtes du signal), ce qui nous a permis de garantir l'imperceptibilité tout en étant robuste aux attaques de compression et au filtrage. L'utilisation de transformée de Fourier rend le message plus robuste à la compression, puisqu'elle utilise le même espace que celui utilisé pour le codage. De plus, ces schémas peuvent offrir une plus grande robustesse aux

attaques par filtrage, puisque les coefficients tatoués sont répartis dans tout le fichier audio. Le fait de modifier le moins de valeurs possible nous a permis de générer moins de distorsions, ce qui nous donne des résultats SNR raisonnables (plus de 36dB).

Après avoir comparé nos approches aux travaux récents, nous pouvons conclure que notre deuxième variante offre de bons résultats d'imperceptibilité pour l'insertion dans les coefficients DFT et FrFT. Et la première variante est plus adaptée à l'insertion dans les coefficients QDFT. Les résultats obtenus restent satisfaisants par rapport à certains travaux récents avec un SNR supérieur à 36,90dB en moyenne. Cependant, en utilisant des fichiers audio de petite taille (ce qui réduit le nombre de coefficients) et puisque notre approche nécessite trois coefficients pour insérer deux bits avec la première variante et deux coefficients pour insérer un bit avec la deuxième variante, la capacité de nos approches est très limitée dans le domaine fréquentiel (entre 9002 et 17937 bits par fichier).

## **5.3 APPROCHES SPATIALES ET FREQUENTIELLES POUR LA PROTECTION DES FICHIERS AUDIO**

### **5.3.1 Contexte d'étude et problématique**

Actuellement, il est devenu très facile d'obtenir, de reproduire et de distribuer du contenu numérique sans perte de qualité [12], ce qui est devenu une préoccupation majeure pour les industries d'édition de contenu multimédia. Le transfert de données sensibles ne peut se faire avec de tels risques et doit donc être sécurisé. Plusieurs méthodes ont été déployées pour pallier à ce problème [211]. La cryptographie a été l'outil efficace utilisé pour protéger et sécuriser les informations en transmettant des données chiffrées. Elle offre ainsi des solutions pour protéger la confidentialité des données, pour assurer leur intégrité ou pour connaître l'identité de la personne qui les envoie [14]. Mais ces protections n'agissent que pendant la transmission et la diffusion. Une fois que les données sont en clair, elles ne contiennent plus aucune protection [4].

Pour surmonter ce problème, des techniques de dissimulation de données telles que la stéganographie ou le tatouage permettent de dissimuler des informations dans un fichier et de maintenir ainsi le fichier protégé même après réception [16]. Le tatouage est la méthode la plus pratique pour aborder la question de la confidentialité et du droit d'auteur et pour garantir l'authenticité des produits multimédias [17]. L'idée de base du tatouage numérique est de créer des métadonnées contenant des informations sur le contenu numérique, puis de cacher ces métadonnées dans ce contenu [18]. Les informations stockées sous forme de métadonnées peuvent avoir différents formats, comme une chaîne de caractères ou une image binaire, voire une image numérique. Un système de tatouage efficace doit garantir l'imperceptibilité des données cachées [19]. L'imperceptibilité, la robustesse et la capacité de la marque sont les propriétés importantes qui caractérisent la technique de tatouage.

La robustesse garantit que la marque inaudible ne peut être détruite sans que le support ne soit gravement dégradé [16]. La transparence ou l'imperceptibilité indique que, même lors de l'insertion d'informations secrètes, le document tatoué n'est perceptivement pas perturbé. La capacité indique la quantité d'informations pouvant

être insérée sans être audible [20]. Ces propriétés sont étroitement liées mais ont l'effet inverse. Si la capacité est grande, elle a un effet positif sur la robustesse mais un effet négatif sur l'imperceptibilité et vice versa.

Dans notre travail, nous proposerons un nouveau schéma de tatouage pour les fichiers audio. Notre objectif est de trouver un compromis entre la capacité et l'imperceptibilité afin de cacher autant de données que possible tout en minimisant la dégradation du fichier. Notre première expérience consiste à appliquer notre schéma de tatouage dans le domaine spatial, le processus de dissimulation consistera à échanger les trois valeurs d'échantillons consécutifs en fonction des bits de la marque. Ensuite, la deuxième expérience consistera à appliquer notre schéma de tatouage dans le domaine fréquentiel, les coefficients DCT du fichier seront calculés et l'échange sera effectué sur les valeurs de ces coefficients. Notre dernière expérience consistera à appliquer notre schéma dans le domaine de la multi-résolution. Pour cela, une DWT sera appliquée au conteneur et les données seront insérées en échangeant les coefficients obtenus par DWT.

Pour chaque domaine d'insertion, les mesures de distorsion seront calculées afin d'avoir une idée plus objective des distorsions qui se sont produites dans le conteneur. Cela permettra également de déterminer le taux d'imperceptibilité obtenu pour chaque domaine. En outre, la capacité de dissimulation varie en fonction du domaine d'insertion. Nous pourrions donc avec un état récapitulatif (capacité/imperceptibilité) déterminer, pour quel domaine d'insertion, le schéma proposé est utilisable. Cette deuxième partie de nos contributions est organisée comme suit, La section 2 sera consacrée à la description et à la présentation de notre schéma proposé ainsi qu'à leur fonctionnement pour chaque domaine d'insertion. Les expériences et l'évaluation des performances de chaque schéma seront présentées dans la section 3. L'analyse des résultats et les perspectives seront discutées dans la section 4.

### **5.3.2 Approches proposées pour le tatouage aveugle**

Le processus d'insertion proposé dans ce travail consiste à réordonner les valeurs du fichier en fonction des bits de la marque. Comme les valeurs adjacentes sont généralement similaires, une permutation de deux valeurs voisines peut ne pas entraîner une dégradation significative du fichier. L'insertion dans le domaine spatial ne nécessite pas de transformation du fichier, La marque est insérée directement en utilisant les valeurs des échantillons du fichier. Pour l'insertion dans le domaine fréquentiel, une DCT est d'abord appliquée, puis les coefficients de la bande de moyenne fréquence sont utilisés pour le processus d'insertion. Pour l'insertion dans le domaine multi-résolution, une DWT à un seul niveau est appliquée afin d'obtenir les composantes d'approximation et de détail. Une permutation des composantes d'approximation sera effectuée pour insérer la marque.

#### **5.3.2.1 Domaine spatial**

La méthode d'insertion proposée consiste à comparer trois valeurs d'échantillon consécutives, cette comparaison générera un élément minimum, un maximum et une moyenne. Les bits de la marque seront regroupés par deux, ce qui constitue quatre combinaisons possibles (00, 01, 10 et 11). Les valeurs (minimum, maximum et moyenne) seront ensuite redistribuées en fonction des bits de la marque à insérer,

comme indiqué dans le tableau 5.6. Dans le schéma de tatouage proposé, trois valeurs d'échantillon sont utilisées pour cacher deux bits. Soit Min, Max et Avr trois valeurs correspondant aux valeurs minimale, maximale et moyenne de trois valeurs d'échantillon successives.

Tableau 5.6 - Condition de redistribution des valeurs de l'échantillon

Réorganisation des valeurs de l'échantillon	Bits de la marque
Min, Avr, Max	0, 0
Max, Avr, Min	1, 1
Min, Max, Avr	0, 1
Max, Min, Avr	1, 0

Pour le processus d'extraction, les valeurs d'échantillon du fichier tatoué sont groupées par trois, puis ces valeurs sont comparées pour déterminer la valeur minimale, maximale et moyenne. L'ordre obtenu permet d'extraire deux bits selon les règles du tableau 5.6.

### 5.3.2.2 Domaine fréquentiel

Dans l'approche proposée pour le domaine fréquentiel, une DCT est appliquée aux trames audio, après l'étape de seuillage et de quantification, la marque est insérée dans les coefficients DCT pour obtenir le fichier tatoué.

#### 1) Utilisation de la transformée en cosinus discrète (DCT)

La DCT est une technique de transformation audio qui transforme les données du domaine spatial en un domaine de transformation [212]. Cette transformation linéaire transforme une matrice de n éléments en une autre matrice de n coefficients sous la forme d'une somme cumulée de fonctions cosinus de différentes fréquences. La matrice est ainsi divisée en trois bandes de fréquences : basse fréquence (LF), moyenne fréquence (MF) et haute fréquence (HF). La plus grande quantité d'énergie est concentrée dans la bande LF, tandis que la plus petite quantité d'énergie est concentrée dans la bande HF. L'insertion de la marque dans la bande LF peut affecter l'audibilité du signal, tandis que l'insertion de la marque dans la bande HF peut affecter la qualité des bits tatoués.

Par conséquent, dans notre approche, la marque est insérée dans la bande MF car elle est la moins sensible au bruit. Après le calcul d'une DCT par la formule (5.1), un seuillage est effectué pour augmenter l'efficacité de la compression. Si les valeurs absolues des coefficients (non nuls) de la matrice DCT obtenue sont inférieures à un certain seuil, ces coefficients seront éliminés (mis à zéro).

$$F_{u,v} = \frac{1}{\sqrt{2N}} C_u \sum_{x=1}^N f(x) \cdot \cos\left[\frac{(2x+1)u\pi}{2N}\right] \tag{5.1}$$

$$\text{où } C_u = \begin{cases} \frac{1}{\sqrt{2N}} & \text{if } u = 0 \\ 1 & \text{if } u > 0 \end{cases}$$

### ***Processus de quantification***

Une quantification des coefficients DCT obtenus est effectuée. La quantification de chaque bloc regroupe les ensembles de valeurs similaires. Cela permet de minimiser les données nécessaires pour représenter le fichier audio. Une seule valeur de quantification représentera donc une gamme de valeurs [213]. Afin de calculer la matrice de quantification  $d_q(i, j)$ , les valeurs maximales  $d_1(i, j)$  et les valeurs minimales  $d_2(i, j)$  sont déterminées. Ensuite, la matrice de quantification est calculée par la formule suivante :

$$d_q(i, j) = \frac{(-1 + 2^Q)(d(i, j) - d_1(i, j))}{d_1(i, j) - d_2(i, j)} \quad (5.2)$$

Où  $Q$  représente le nombre de bits nécessaires pour coder chaque élément de la matrice  $d(i, j)$ .

### ***Processus d'insertion***

Pour le processus d'insertion, trois coefficients successifs sont utilisés et comparés pour générer le minimum, la moyenne et le maximum. Les valeurs (minimum, maximum et moyenne) seront ensuite redistribuées en fonction des bits de la marque à insérer, comme indiqué dans le tableau 5.7. Dans ce tableau, Min, Max et Avr sont trois valeurs correspondant aux valeurs minimale, maximale et moyenne de trois coefficients DCT successifs.

Tableau 5.7 - Condition de redistribution des coefficients DCT

Réorganisation des coefficients DCT	Bits de la marque
DCT-Min, DCT-Avr, DCT-Max	0, 0
DCT-Max, DCT-Avr, DCT-Min	1, 1
DCT-Min, DCT-Max, DCT-Avr	0, 1
DCT-Max, DCT-Min, DCT-Avr	1, 0

Pour le processus d'extraction, une DCT est appliquée aux trames audio, après l'étape de seuillage et de quantification, les coefficients obtenus sont groupés par trois. Ces valeurs sont ensuite comparées afin de déterminer la valeur minimale, maximale et moyenne. L'ordre résultant permet d'extraire deux bits selon les règles du tableau 5.7.

### **5.3.2.3 Domaine Multi-résolution DWT**

Dans l'approche proposée pour le domaine multi-résolution, une DWT à un seul niveau est calculée, puis les bits du message secret sont intégrés dans les coefficients obtenus. Avant d'insérer la marque, une DWT à un seul niveau est appliquée en utilisant le filtre d'échelle (passe-bas) LoD et le filtre d'ondelette (passe-haut) HiD, comme le montre la figure 5.5.



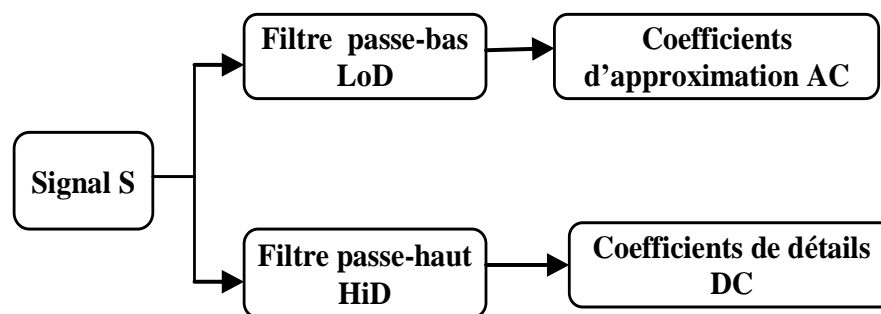


Fig. 5.5 - Schéma DWT à un niveau

La transformation en ondelettes décompose le signal audio en composantes d'approximation et de détail [214]. Les coefficients d'approximation sont des composantes de basse fréquence et les coefficients de détail sont des composantes de haute fréquence [204]. Après avoir obtenu les composantes AC et DC, la substitution est effectuée dans des coefficients AC obtenus. Un minimum, un maximum et une moyenne de trois composantes successives sont calculés. La permutation est ensuite effectuée en fonction des bits de la marque comme indiqué dans le tableau 5.8. Une IDWT est ensuite appliquée pour générer le fichier tatoué.

Tableau 5.8 - Condition de redistribution des coefficients AC

Réorganisation des coefficients AC	Bits de la marque
AC-Min, AC-Avr, AC-Max	0, 0
AC-Max, AC-Avr, AC-Min	1, 1
AC-Min, AC-Max, AC-Avr	0, 1
AC-Max, AC-Min, AC-Avr	1, 0

Pour le processus d'extraction, une DWT à un seul niveau est calculée ; les coefficients AC obtenus sont groupés par trois. Ces valeurs sont ensuite ordonnées pour déterminer la valeur minimale, maximale et moyenne, l'ordre obtenu permet d'extraire deux bits selon les règles du tableau 5.8.

### 5.3.3 Résultats expérimentaux

Dans cette section, les techniques de tatouage proposées sont testées sur une grande base de données de fichiers audio. Cette base de données (BBC Sound Effects) [215] contient plus de 16.000 effets sonores. Une description complète de la base donnée des effets sonores de BBC est présentée dans l'annexe 2 de cette thèse. Nous avons appliqué notre approche dans les trois domaines d'insertion (spatial, fréquentiel et multi-résolution).

#### 5.3.3.1 Mesures de la performance

L'évaluation de la qualité peut se faire de deux façons, soit de manière subjective, soit de manière objective. Les méthodes basées sur une évaluation subjective nécessitent un effort humain. En effet, un humain est nécessaire pour écouter les

fichiers audio et évaluer leurs qualités. Ces méthodes sont coûteuses et l'évaluation peut varier selon l'évaluateur. Les méthodes d'évaluation objectives sont basées sur des formules mathématiques pour quantifier la qualité des fichiers audio, ces méthodes ne sont pas coûteuses et sont stables et fiables.

Pour vérifier le critère de transparence et d'imperceptibilité, le fichier audio tatoué doit être de la même qualité que l'original, le fichier tatoué et l'original doivent être similaires. Les mesures de qualité sont basées sur le calcul de la proximité du fichier tatoué par rapport au fichier original. Afin de mesurer et de tester les performances des approches proposées, la mesure SNR (Signal-to-noise ratio) est utilisée. Le SNR est une métrique qui mesure la quantité de bruit ajoutée au fichier audio pendant le processus de tatouage [216]. Il est calculé comme suit :

$$SNR = \frac{\sum_1^M \sum_1^N I^2}{\sum_1^M \sum_1^N (I_w - I)^2} \quad (5.3)$$

Où  $I$  est la valeur de l'échantillon du fichier original et  $I_w$  celle du fichier tatoué.

Une valeur élevée du SNR est une caractéristique souhaitable et signifie que le niveau de distorsion est faible, le SNR est l'évaluation la plus couramment utilisée [217]. Afin d'effectuer une meilleure évaluation de chaque algorithme, nous avons également calculé l'erreur quadratique moyenne (Mean Squared Error –MSE), Le MSE fournit des informations sur la dégradation ou la distorsion introduite. Plus la valeur MSE est faible, plus le fichier tatoué est apprécié. Elle est définie par la formule suivante :

$$MSE = \frac{1}{M*N} \sum_i \sum_j (I(i,j) - J(i,j))^2 \quad (5.4)$$

Où  $I(i,j)$  est la valeur de l'échantillon du fichier original et  $J(i,j)$  celle du fichier tatoué, les deux fichiers étant de taille  $[M \times N]$ .

SSIM (Structural SIMilarity) est une mesure de similarité entre deux fichiers. L'idée de SSIM est de mesurer la similarité de structure entre les deux fichiers, plutôt qu'une différence de valeur d'échantillon à valeur d'échantillon comme le fait par exemple le SNR. La métrique SSIM est calculée sur plusieurs fenêtres d'un fichier audio [218]. La mesure entre deux fenêtres  $x$  et  $y$  de taille  $N \times N$  est :

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\delta_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\delta_x^2 + \delta_y^2 + C_2)} \quad (5.5)$$

Où  $\mu_x$  - moyenne de  $x$ ,  $\mu_y$  - moyenne de  $y$ ,  $\delta_x^2$  - variance de  $x$ ,  $\delta_y^2$  - variance de  $y$ ,  $\delta_{xy}$  - covariance de  $x$  et  $y$ ,  $C_1 = k_1 L_1$  - constante pour éviter l'instabilité lorsque  $\mu_x^2 + \mu_y^2$  est proche de zéro;  $k_1 = 0.01$ ,  $L_1 = 255$ , et  $C_2 = k_2 L_2$  - constante pour éviter l'instabilité lorsque  $\delta_x^2 + \delta_y^2$  est proche de zéro;  $k_2 = 0.03$ ,  $L_2 = 255$ .

Afin d'effectuer une meilleure évaluation de chaque algorithme, la variance (VAR), l'écart-type (STD) et le taux d'erreur binaire (BER) ont également été calculés. Pour évaluer la robustesse du processus d'intégration, la corrélation croisée normalisée (*normalized cross correlation*-NCC) est également utilisée en comparant la marque originale insérée et la marque extraite.

### 5.3.3.2 Application de nos approches dans le domaine spatial

Afin d'expérimenter nos approches dans le domaine spatial, les fichiers audio de notre base de données ont été tatoués, une séquence aléatoire de bits a été insérée en substituant la position des valeurs d'échantillon selon le tableau 5.6.

Comme on peut le voir sur la figure 5.6, les distorsions générées sont visuellement indétectables. Afin d'analyser les résultats obtenus, les mesures de similarité et de distorsion ont été calculées dans le "Tableau 5.9" en comparant les fichiers originaux et tatoués pour différents taux d'insertion.

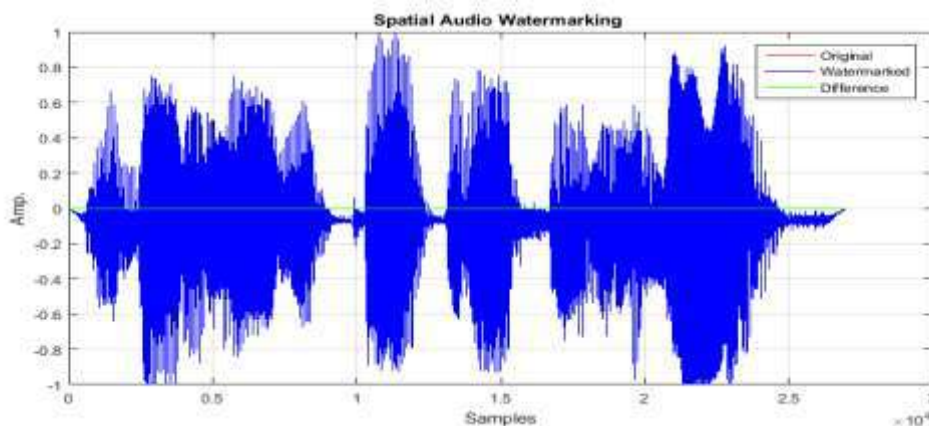


Fig. 5.6 - Différence entre le fichier audio original et le fichier tatoué en LSB

Tableau 5.9 - Résultats d'imperceptibilité pour l'insertion dans le domaine spatial

Taux d'insertion	10%	20%	30%	40%	50%
Capacité (Bits)	2696	5392	8088	10784	13480
SNR	89,72	79,42	42,19	35,22	31,06
MSE	0	0,0003	1	2,7647	5,3539
SSIM	1	1	0,9987	0,9786	0,9601
VAR	0,0938	0,0984	0,3459	0,1827	0,2121
STD	0,2166	0,2566	0,2261	0,4212	0,4587
BER	0,0018	0,0026	0,3459	0,9407	1,1453

### 5.3.3.3 Application de nos approches dans le domaine fréquentiel

Afin d'expérimenter nos approches dans le domaine fréquentiel, une DCT a été calculée pour chaque fichier de notre base de données (après division en blocs), puis une séquence aléatoire de bits a été intégrée en réarrangeant les coefficients DCT quantifiés obtenus pour chaque fichier "Figure 5.7". Les mesures de distorsion sont calculées entre les fichiers originaux et les fichiers tatoués " Tableau 5.10".

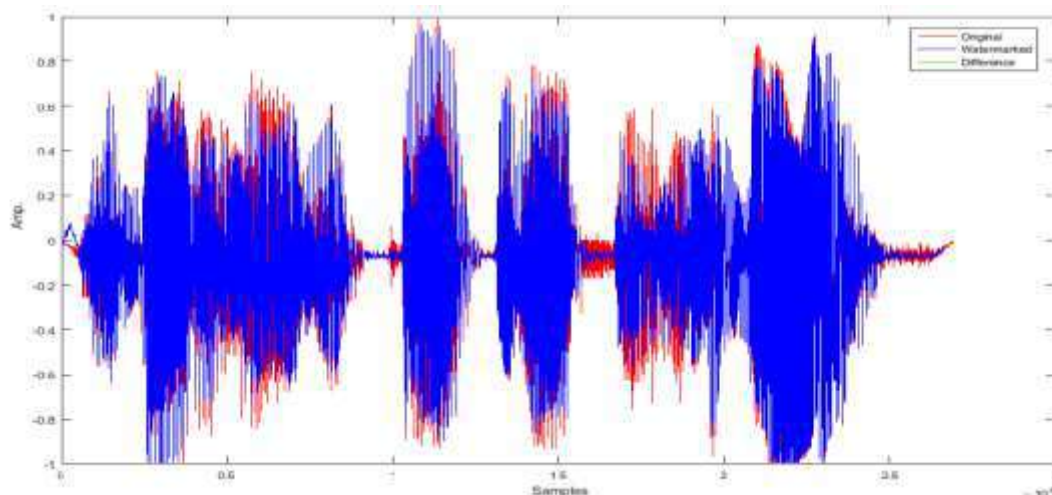


Fig. 5.7 - Différence entre le fichier audio original et le fichier tatoué en DCT

Tableau 5.10 - Résultats d'imperceptibilité pour l'insertion dans le domaine de la fréquence

Taux d'insertion	10%	20%	30%	40%	50%
Capacité (Bits)	665	1330	1995	2660	3325
SNR	32,553	32,629	33,077	33,2988	33,7678
MSE	0,0054	0,0075	0,0293	0,0574	0,0821
SSIM	0,7437	0,7182	0,0026	0,4963	0,3498
VAR	0,1876	0,1911	0,636	0,2435	0,25
STD	0,4332	0,4371	0,216	0,4934	0,5
BER	0,0027	0,0039	0,5188	1,4111	1,7180

### 5.3.3.4 Application de nos approches dans le domaine de la multi-résolution

Afin d'expérimenter nos approches dans le domaine multi-résolution, chaque fichier de notre base de données est décomposé en sous-bandes (AC et DC) en appliquant une DWT à un seul niveau. Les coefficients AC seront utilisés pour insérer la marque. Après avoir dissimulé notre message qui est une séquence aléatoire de 90 000 bits, les fichiers obtenus ne permettent pas visuellement de détecter des distorsions "Figure 5.8".

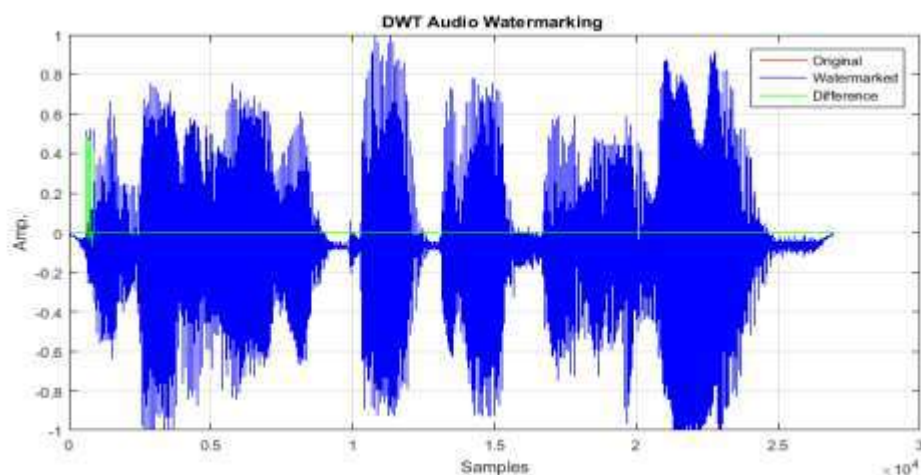


Fig. 5.8 - Différence entre le fichier audio original et le fichier tatoué en DWT

Afin de mesurer les performances et évaluer les variantes utilisées, des mesures de similarité et de distorsions entre les fichiers originaux et tatoués ont été calculées. Les valeurs moyennes obtenues sont présentées dans le tableau 5.11.

Tableau 5.11 - Résultats d'imperceptibilité pour l'insertion dans le domaine multi-résolution

Taux d'insertion	10%	20%	30%	40%	50%
Capacité(Bits)	1348	2696	4044	5392	6740
SNR	75,8449	31,2419	24,8762	21,4494	20,3424
MSE	0	0,0008	0,0033	0,0072	0,0092
SSIM	1	0,9837	0,9348	0,8662	0,8321
VAR	0	0,0058	0,0558	0,1219	0,1742
STD	0	0,0761	0,2363	0,3491	0,4174
BER	0,0023	0,0033	0,4323	1,1759	1,4317

### 5.3.3.5 Comparaison des résultats

La figure 5.9 montre les valeurs de distorsion obtenues par les trois algorithmes de tatouage à différents niveaux d'insertion. Comme nous pouvons le constater, notre approche dans le domaine spatial et multi résolution offre de très bons résultats contrairement au domaine fréquentiel pour des insertions inférieures à 20% de la taille du fichier. Cependant, pour des insertions supérieures à 20%, les distorsions augmentent considérablement pour les insertions dans le domaine spatial et multi résolution. Contrairement au domaine fréquentiel où le SNR reste stable pour tous les niveaux d'insertion. Dans le domaine spatial une combinaison entre les échantillons de valeurs de données nous a permis de cacher deux bits en modifiant une seule valeur. Modifier le moins de valeurs possible nous a permis de générer moins de distorsions, ce qui nous donne des résultats SNR raisonnables qui sont supérieurs à 59dB.

Pour l'insertion en multi-résolution, la décomposition en sous-bandes nous a permis d'affiner l'isolation des composantes basses fréquences, c'est-à-dire un espace d'insertion moins sensible tout en conservant le contenu spatial du fichier. Les tests

d'imperceptibilité révèlent des niveaux de SNR acceptables et satisfaisants par rapport à certains travaux récents avec un SNR supérieur à 52dB en moyenne.

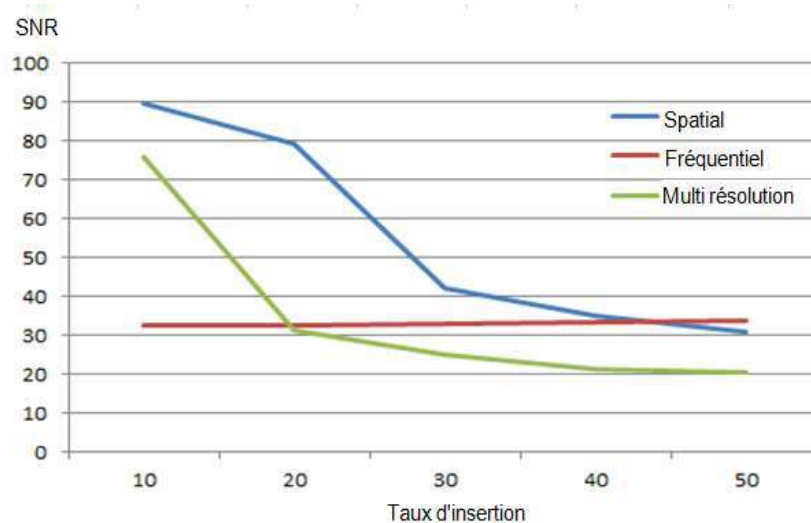


Fig. 5.9 - SNR obtenu pour les trois domaines d'insertion

Afin de comparer et de situer nos approches par rapport aux travaux récents, pour chaque domaine d'insertion, nous avons comparé les résultats de distorsion obtenus avec ceux des travaux connexes présentés dans la dernière section du troisième chapitre de cette thèse, intitulé "étude de l'état de l'art sur le tatouage audio numérique".

Tableau 5.12 - Comparaison de l'imperceptibilité avec des travaux connexes

Domaine spatial	LSB1	Proposée	Hosny & al.[128]	Kundu & Kaur [129]	Devi & Pugazhenthib [130]	
SNR	68,63	59,52808	67,63	79,73	73,6	
Domaine fréquentiel	LSB1	Proposée	Irawati & al.[138]	Jayarani & al.[137]	Karajeh & Maqableh [139]	Xue & al.[140]
SNR	32,39	33,06512	26,39	29,15	35,67	24,3813
Domaine multi-résolution	LSB1	Proposée	Saadi & al.[142]	Kaur & Dutta [143]	Jiang & al.[141]	Hu & al.[144]
SNR	39,72	52,75096	41,6001	42,419	40,6	24,63

Comme nous pouvons le voir dans le tableau 5.12, les valeurs de SNR obtenues par nos approches dans les trois domaines d'insertion restent raisonnables par rapport aux travaux connexes. Dans la première approche, le processus de dissimulation combine trois valeurs pour intégrer deux bits, avec la possibilité de modifier une seule valeur. De ces trois valeurs, une seule peut être modifiée, ce qui réduit la probabilité de modification contrairement aux autres approches. Cela implique moins de modification et donc moins de distorsion du fichier hôte, cela explique le bon taux de SNR obtenu et donc une imperceptibilité raisonnable. Cependant, comme notre première approche nécessite trois valeurs pour insérer deux bits, cela a un impact sur la capacité de l'approche. Dans le domaine spatial, cela peut être négligeable puisque tous les échantillons de données du fichier audio sont AC-disponibles pour l'insertion. Mais cela devient plus problématique dans l'insertion du domaine fréquentiel où le nombre de coefficients DCT quantifiés est réduit.

### 5.3.3.6 Test de robustesse

Pour évaluer la robustesse des schémas proposés, plusieurs attaques ont été exécutées sur les échantillons tatoués, et les résultats obtenus après l'extraction de la marque sont comparés aux travaux connexes. La corrélation normalisée (NC) est utilisée pour évaluer la robustesse du schéma de tatouage en comparant la marque originale et la marque extraite après l'attaque des échantillons tatoués. Les résultats obtenus sont présentés dans le tableau 5.13. Normalement, une  $NC > 0,85$  implique une similarité considérable entre la marque originale et la marque extraite [219].

Tableau 5.13 - Comparaison de la robustesse avec les travaux connexes

Attaque	Domaine spatial		Domaine fréquentiel		Domaine multi-résolution	
	Proposée	Karajeh & Maqableh [139]	Proposée	Kaur & Dutta [143]	Proposée	
Bruit gaussien	0.9873	0.9715	0.9830	0.9930	0.9978	
Ré-échantillonnage	0.9778	0.9619	0.9858	0.9969	0.9963	
Filtre passe-bas	0.9406	0.9708	0.9446	0.9429	0.98521	
Filtre passe-haut	0.9630	0.9706	0.9686	0.9068	0.9011	
Compression MP3	0.9746	0.9720	0.9894	0.9348	0.94325	
Recadrage (Cropping)	0.9887	0.9558	0.9575	0.9998	0.9963	

Comme nous pouvons le voir dans le tableau 5.13, les schémas proposés génèrent des échantillons tatoués raisonnablement robustes contre diverses attaques avec une marque de haute qualité. Habituellement, dans le domaine spatial, si les LSB sont modifiés, l'extraction de la marque devient impossible. Cependant, dans les schémas proposés, nous pouvons récupérer la marque incorporée même après la distorsion des bits LSB. Ceci s'explique par le peu de modification effectuée dans les approches proposées, ce qui réduit la probabilité de modification. Nous pouvons donc conclure des résultats obtenus que nos approches sont résistantes à plusieurs attaques malveillantes de suppression de la maque en garantissant une valeur SNR raisonnable.

### 5.3.4 Analyse et discussion des résultats

Dans cette contribution, une nouvelle approche a été proposée pour le tatouage des fichiers audio, cette approche a été appliquée dans trois domaines d'insertion. Dans le domaine spatial, une combinaison entre les valeurs d'échantillons nous a permis de cacher deux bits en modifiant une seule valeur. Modifier le moins de valeurs possible a permis de générer moins de distorsions, ce qui donne des résultats SNR raisonnables, supérieurs à 59dB.

Notre approche a ensuite été appliquée dans le domaine de transformation DCT (fréquentiel). L'utilisation de transformée rend le message plus robuste à la compression, puisqu'elle utilise le même espace que celui utilisé pour le codage. En plus, ces schémas peuvent offrir une meilleure robustesse contre les attaques de filtrage, car les coefficients marqués sont répartis dans tout le fichier audio.

Enfin, les approches proposées ont été appliquées au domaine multi-résolution DWT en substituant les coefficients des sous-bandes AC obtenus. La décomposition en sous-bandes nous a permis d'affiner l'isolation des composantes basses fréquences,

ce qui constitue un espace d'insertion moins sensible tout en conservant le contenu spatial du fichier.

Pour les travaux futurs, des améliorations doivent être apportées afin d'améliorer notre approche de l'insertion dans le domaine fréquentiel, le but étant de réduire les distorsions générées lors de l'insertion et d'augmenter la capacité de dissimulation qui est un critère essentiel pour une méthode de tatouage.

## **5.4 UN SCHEMA DE TATOUAGE BASE SUR LA COMBINAISON DE VALEUR DE PARITE POUR LA PROTECTION DES SONS DE BATTEMENTS DE CŒUR**

### **5.4.1 Contexte d'étude et problématique**

Afin d'améliorer la sécurité des données échangées en télémédecine, nous proposons dans ce travail un schéma de tatouage aveugle pour la protection des sons du rythme cardiaque. La marque est composée des informations du patient et des données d'acquisition. Ainsi, lors du processus d'extraction, une comparaison entre les informations du patient et la marque indiquera si les données ont été modifiées. Cela permettra de s'assurer que l'échantillon correspond bien au patient indiqué dans l'enregistrement.

Dans ce travail, nous proposerons deux schémas de tatouage pour les fichiers audio ; notre objectif est de trouver un compromis entre capacité et imperceptibilité afin de cacher un maximum de données tout en minimisant la dégradation du fichier. Notre première expérience sera d'appliquer nos schémas de tatouage dans le domaine spatial, pour cela, les données à cacher seront substituées aux valeurs des échantillons du fichier selon les variantes utilisées. Ensuite, la deuxième expérience sera d'appliquer les deux schémas de tatouage dans le domaine fréquentiel, les coefficients DCT du fichier seront calculés, et la substitution sera effectuée dans les valeurs de ces coefficients. Notre dernière expérience sera d'appliquer nos quatre variantes d'insertion dans le domaine multi-résolution, pour cela, une DWT sera appliquée au conteneur et les données seront insérées en substituant le bit le moins significatif des coefficients obtenus. Pour chaque domaine d'insertion et pour chaque variante, les mesures de distorsion seront calculées (entre le fichier original et le fichier Tatoué) afin d'avoir une idée plus objective des distorsions qui se sont produites dans le conteneur. Cela permettra également de déterminer le taux d'imperceptibilité obtenu pour chaque variante. Par ailleurs, la capacité de dissimulation des deux schémas varie en fonction du domaine d'insertion. Nous pourrions donc avec un état récapitulatif (capacité/imperceptibilité) déterminer, pour chaque domaine d'insertion, la variante la plus adaptée qui offre le meilleur compromis entre capacité et imperceptibilité.

Cette troisième partie de nos contributions est organisée comme suit. La section 1 sera consacrée à la description et à la présentation de nos deux schémas proposés ainsi qu'à leur fonctionnement pour chaque domaine d'insertion. Les expériences et l'évaluation des performances de chaque schéma seront présentées dans la section 2. L'analyse et la discussion des résultats de ce travail et les perspectives seront discutées dans la section 3.



## 5.4.2 Approches proposées

### 5.4.2.1 Domaine spatial

Le processus de substitution proposé dans cette approche consiste à combiner la parité des valeurs d'échantillon. L'idée est de modifier les bits les moins significatifs afin de générer le moins de modifications possibles pour maintenir une imperceptibilité acceptable. Dans notre approche, la marque se présente sous forme binaire et consiste donc en une séquence de 0 et de 1. Ces données nécessaires sont substituées au LSB des valeurs d'échantillon en fonction de la variante spécifique correctement utilisée. Dans le premier schéma de tatouage proposé, trois valeurs d'échantillon sont correctement utilisées pour cacher soigneusement deux bits possibles. Soient R1, R2 et R3 trois valeurs correspondant aux trois valeurs d'échantillon successives et X, Y deux bits du message secret. Pour la deuxième variante, les valeurs A, B et C correspondent aux derniers bits significatifs des trois valeurs d'échantillon successives R1, R2 et R3. Dans le deuxième schéma, deux valeurs d'échantillon sont utilisées pour cacher un bit. La différence entre deux valeurs d'échantillon est calculée. Soit R1 et R2 deux valeurs correspondant à deux valeurs d'échantillon successives et X le bit à intégrer. Pour la deuxième variante, les valeurs A et B correspondent aux derniers bits significatifs des valeurs d'échantillon successives R1 et R2.

Tableau 5.14 - Premier schéma d'insertion proposé

Variante 1		Variante 2	
Condition	Action	Condition	Action
$X = (R1 - R2) \% 2$ $Y = (R1 - R3) \% 2$	Aucun changement requis	$X = (A \oplus B) \% 2$ $Y = (A \oplus C) \% 2$	Aucun changement requis
$X \neq (R1 - R2) \% 2$ $Y = (R1 - R3) \% 2$	Changer la composante R2	$X \neq (A \oplus B) \% 2$ $Y = (A \oplus C) \% 2$	Changer la composante R2
$X = (R1 - R2) \% 2$ $Y \neq (R1 - R3) \% 2$	Changer la composante R3	$X = (A \oplus B) \% 2$ $Y \neq (A \oplus C) \% 2$	Changer la composante R3
$X \neq (R1 - R2) \% 2$ $Y \neq (R1 - R3) \% 2$	Changer la composante R1	$X \neq (A \oplus B) \% 2$ $Y \neq (A \oplus C) \% 2$	Changer la composante R1

Tableau 5.15 - Deuxième schéma d'insertion proposé

Variante 1		Variante 2	
Condition	Action	Condition	Action
$X = (R1 - R2) \% 2$	Aucun changement requis	$X = (A \oplus B) \% 2$	Aucun changement requis
$X \neq (R1 - R2) \% 2$	Change la composante R1	$X \neq (A \oplus B) \% 2$	Change la composante R1

### 5.4.2.2 Domaine fréquentiel

Dans l'approche proposée pour le domaine fréquentiel, une DCT est soigneusement appliquée aux trames audio, après l'étape de seuillage et de quantification, la marque est soigneusement insérée dans les coefficients complexes de la DCT pour obtenir le fichier tatoué.

### 1) *La transformée en cosinus discrète (DCT)*

La DCT est une technique de transformation audio [212] qui transforme les données du domaine spatial en un domaine de transformation. Cette transformation linéaire transforme une matrice de  $n$  éléments en une autre matrice de  $n$  coefficients sous forme de somme cumulée de fonctions cosinus de différentes fréquences. La matrice est ainsi divisée en trois bandes de fréquences : basse fréquence (LF), moyenne fréquence (MF) et haute fréquence (HF). La plus grande quantité d'énergie est concentrée dans la bande LF, tandis que la plus petite quantité d'énergie est concentrée dans la bande HF. L'intégration de la marque dans la bande LF peut affecter l'audibilité du signal, tandis que l'intégration de la marque dans la bande HF peut affecter la qualité des bits tatoués. Par conséquent, dans notre approche, la marque est insérée dans la bande MF car elle est la moins sensible au bruit. Après le calcul d'une DCT par la formule (5.3), un seuillage est effectué pour augmenter l'efficacité de la compression. Si les valeurs absolues des coefficients (non nuls) de la matrice DCT obtenue sont inférieures à un certain seuil, ces coefficients seront éliminés (mis à zéro).

$$F_{u,v} = \frac{1}{\sqrt{2N}} C_u \sum_{x=1}^N f(x) \cdot \cos\left[\frac{(2x+1)u\pi}{2N}\right] \quad (5.3)$$

$$\text{où } C_u = \begin{cases} \frac{1}{\sqrt{2N}} & \text{if } u = 0 \\ 1 & \text{if } u > 0 \end{cases}$$

#### *Processus de quantification*

Une quantification des coefficients DCT obtenus est effectuée. La quantification de chaque bloc regroupe les ensembles de valeurs similaires. Cela permet de minimiser les données nécessaires pour représenter le fichier audio. Une seule valeur quantifiée représentera donc une gamme de valeurs [213]. Pour calculer la matrice de quantification, on choisit la valeur de la quantification  $Q$ , ( $Q$  représente le nombre de bits nécessaires pour coder chaque élément de la matrice DCT). Les deux valeurs MAX et MIN de la matrice DCT (DCTmax, DCTmin) sont déterminées. Ensuite, la matrice DCTQ est calculée par la formule suivante :

$$\text{DCTQ}(i,j) = \frac{(-1 + 2^Q)(\text{DCT}(i,j) - \text{DCTmin})}{\text{DCTmax} - \text{DCTmin}} \quad (5.4)$$

#### *Processus d'insertion*

Dans le premier schéma complexe, la parité de trois coefficients successifs est calculée avec précision, puis comparée aux bits possibles à cacher soigneusement. Dans le cas spécifique de l'inégalité utilisée, la parité d'une constante déterminée est ensuite modifiée pour satisfaire suffisamment l'égalité selon les règles spécifiques du tableau 5.14. Pour la première variante, R1, R2 et R3 représentent typiquement trois coefficients successifs et X, Y deux bits possibles du message privé à cacher efficacement. Pour la deuxième variante développée, les valeurs A, B et C

représentent typiquement les restes possibles de la division entière sur 2 des valeurs possibles R1, R2 et R3. Dans le deuxième schéma proposé, la parité de deux coefficients successifs est calculée avec précision ; la substitution éventuelle d'un bit est effectuée en exécutant correctement l'une des règles établies dans le tableau 5.15. Pour la première variante, R1 et R2 représentent avec précision deux coefficients successifs et X le bit spécifique du message privé à intégrer de manière typique. Pour la deuxième variante, les valeurs A et B représentent les restes de la division entière sur 2 des valeurs R1 et R2.

#### 5.4.2.3 Domaine multi-résolution

Dans l'approche proposée pour le domaine multi-résolution, une DWT est calculée, puis, les bits du message secret sont intégrés dans les coefficients obtenus. Une inverse IDWT est ensuite appliquée pour générer le fichier tatoué. Avant d'intégrer la marque, une DWT est appliquée. La transformation en ondelettes décompose le signal audio en composantes d'approximation et de détail [214]. Les coefficients d'approximation sont des composantes de basse fréquence et les coefficients de détail sont des composantes de haute fréquence [204]. Après avoir obtenu les composantes AC et DC, la substitution est effectuée dans les coefficients des composantes AC obtenues. Dans le premier schéma élaboré, comme pour l'approche fréquentielle, la parité de trois coefficients successifs est soigneusement calculée puis comparée aux éventuels bits à intégrer. Dans le cas exceptionnel d'une inégalité, la parité d'un coefficient complexe est alors modifiée pour satisfaire suffisamment l'égalité fondamentale en acceptant les règles spécifiques du tableau 5.14. Pour la première variante, R1, R2 et R3 représentent typiquement trois constantes successives des éléments nécessaires AC et X, Y deux bits possibles du message privé à cacher soigneusement. Pour la deuxième variante, les valeurs A, B et C représentent les restes de la division entière sur 2 des valeurs R1, R2 et R3. Dans le deuxième schéma proposé, la parité de deux coefficients successifs des éléments AC nécessaires est calculée avec précision ; la substitution éventuelle d'un bit est effectuée en exécutant correctement l'une des règles spécifiques du tableau 5.15. Pour la première variante, R1 et R2 représentent avec précision deux coefficients successifs des éléments essentiels et X le bit possible du message privé à cacher soigneusement. Pour la seconde variante, les valeurs A et B représentent les restes de la division entière sur 2 des valeurs R1 et R2.

#### 5.4.3 Expériences et résultats

Dans cette section, les techniques de tatouage proposées sont testées sur une grande base de données de fichiers de sons de battements cardiaques [220] : Les données ont été recueillies à partir de deux sources : 176 fichiers de patients provenant du grand public via l'application *iPhone iStethoscope Pro* et 176 fichiers de patients provenant d'un essai clinique dans des hôpitaux locaux utilisant correctement le stéthoscope numérique *DigiScope*. Une description complète de la base donnée des sons de battements cardiaques est présentée dans l'annexe 3 de cette thèse.

Nous avons correctement appliqué chaque approche dans les trois domaines d'insertion. Dans ce qui suit, Sch1V1 et Sch1V2 feront référence à la première et à la deuxième variante de la première approche proposée, Sch2V1 et Sch2V2 feront référence à la première et à la deuxième variante de la deuxième approche proposée.

### 5.4.3.1 Mesures de la performance

Pour déterminer et vérifier la validité des méthodes proposées, on peut mesurer le rapport signal/bruit. Le SNR [221] est une mesure classique qui quantifie avec précision la quantité de bruit spécifique généralement introduit dans le signal audio pendant le tatouage. Une valeur SNR élevée est souhaitable et implique que le degré de distorsion est faible; la mesure SNR est la méthode de mesure la plus fréquemment utilisée.

### 5.4.3.2 Application de nos approches dans le domaine spatial

Afin d'expérimenter nos approches dans le domaine spatial, les fichiers audio de notre base de données ont été tatoués, une séquence de bits représentant les informations du patient et les données d'acquisition ont été substituées aux bits les moins significatifs de l'échantillon de données. Dans notre première expérience, seul le dernier bit est modifié. Dans la deuxième expérience, les deux derniers bits sont modifiés, la modification du deuxième bit n'affecte pas la parité des valeurs et les équations du tableau 5.14 et du tableau 5.15 restent valides. Pour la troisième expérience, les trois derniers bits sont modifiés selon les équations du tableau 5.14 et du tableau 5.15.

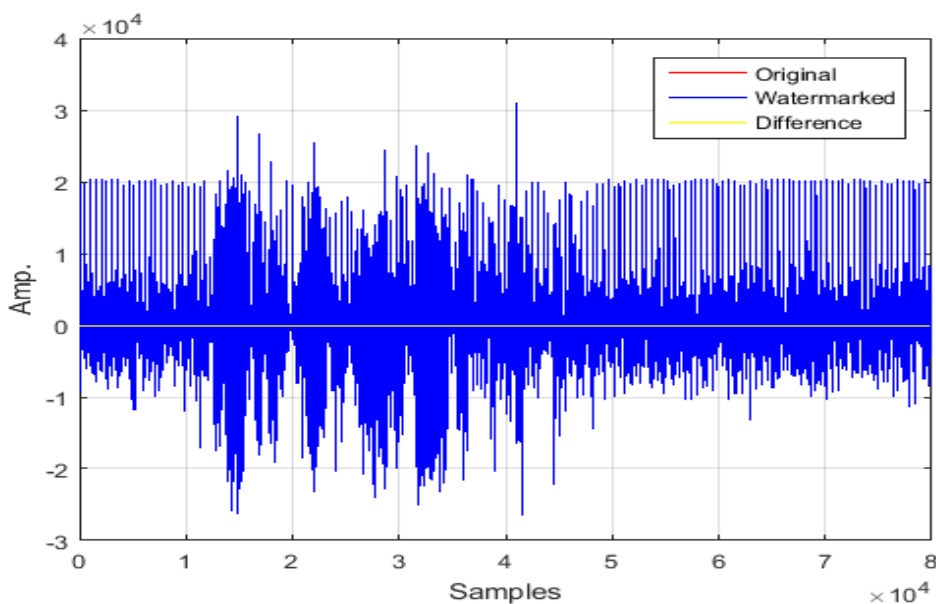


Fig. 5.10 - Différence entre le fichier audio original et le fichier tatoué à l'aide de Sch2.V2

Comme nous pouvons le voir sur la figure 5.10, les distorsions générées sont visuellement indétectables. Afin d'analyser les résultats obtenus "Tableau 5.16" les mesures de similarité et de distorsion ont été calculées en comparant les fichiers originaux et tatoués. Les résultats obtenus par une simple substitution LSB ont également été calculés pour voir la différence entre notre approche et une simple substitution LSB. En effet, une substitution LSB modifie systématiquement toutes les valeurs lors du processus d'intégration, ce qui augmente les distorsions (68,6321dB en moyenne). Notre approche, par contre, ne modifie qu'une valeur sur trois, ce qui réduit la distorsion du fichier tatoué. Comme nous pouvons le voir dans les résultats obtenus dans le "Tableau 5.16", toutes les variantes proposées génèrent moins de

distorsion pendant le processus d'intégration que la méthode LSB, à l'exception de la variante Sch1.V2, ceci s'explique également par le nombre de valeurs modifiées.

Tableau 5.16 - Résultats d'imperceptibilité pour l'insertion dans le domaine spatial

Substitution de 1 bit	15.000 bits	20.000 bits	25.000 bits	30.000 bits	Moyenne
Sch2.V2	83,4430	72,8360	68,4599	63,4304	72,0423
Sch2.V1	81,7935	72,7468	69,3776	62,8172	71,6838
Sch1.V1	81,8480	71,5825	66,1580	60,0870	69,9189
LSB1	80,5659	69,9086	64,7200	59,3338	68,6321
Sch1.V2	73,9006	78,0009	65,1374	50,6372	66,9190
Substitution de 2 bits	15.000 bits	20.000 bits	25.000 bits	30.000 bits	Average
Sch2.V1	77,2472	70,8889	62,9688	58,2630	67,3420
Sch2.V2	77,4098	67,4200	63,7816	58,7989	66,8526
Sch1.V1	76,8983	67,7484	62,8997	57,3613	66,2269
LSB2	74,9906	63,9234	60,5400	54,9046	63,5897
Sch1.V2	78,0023	64,3567	55,1887	53,5005	62,7621
Substitution de 3 bits	15.000 bits	20.000 bits	25.000 bits	30.000 bits	Average
Sch1.V1	72,4349	64,3836	59,3592	53,9526	62,5326
Sch2.V2	75,3138	61,1884	60,9388	52,6061	62,5118
Sch2.V1	70,0062	61,0751	54,7676	49,5317	58,8452
LSB3	69,2433	60,2924	54,2560	51,2694	58,7653
Sch1.V2	69,9212	59,9488	54,9090	44,2055	57,2461

### 5.4.3.3 Application de nos approches dans le domaine fréquentiel

Afin d'expérimenter nos approches dans le domaine fréquentiel, une DCT a été calculée pour chaque fichier de notre base de données (après division en blocs), puis une séquence de bits (représentant les informations sur le patient et les données d'acquisition) a été substituée aux bits les moins significatifs des coefficients DCT quantifiés obtenus pour chaque fichier (Figure 5.11).

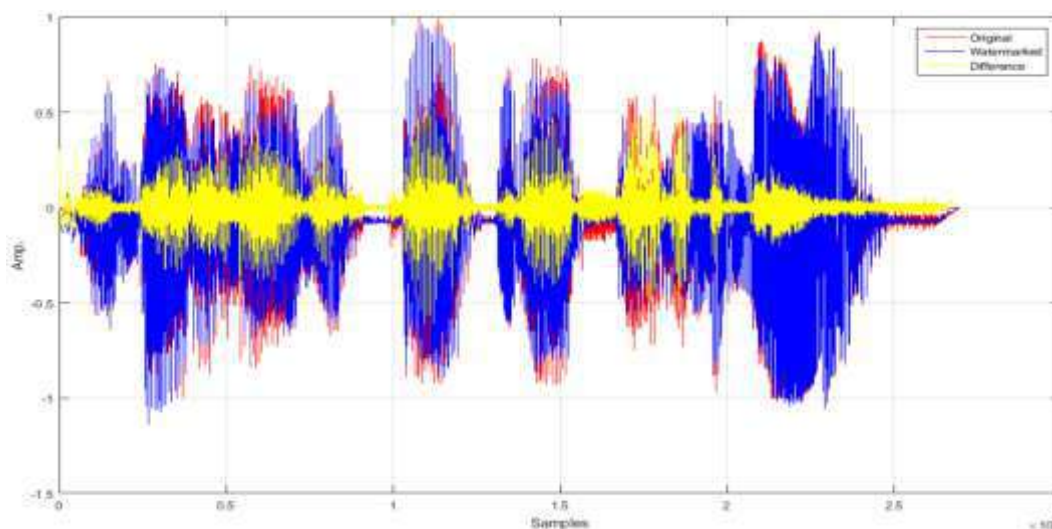


Fig. 5.11 - Différence entre le fichier audio original et le fichier tatoué à l'aide de Sch2.V2

Tableau 5.17 - Résultats d'imperceptibilité pour l'insertion dans le domaine fréquentiel

	2.000 bits	3.000 bits	4.000 bits	Moyenne
Sch2.V2	33,423895	30,1645563	21,7476577	28,4453697
Sch1.V2	32,9794132	29,990219	22,1695192	28,3797171
Sch2.V1	33,1914082	29,4845284	21,9183213	28,1980859
Sch1.V1	32,8679277	29,6838544	21,8455378	28,13244
LSB	32,390956	27,7040684	18,8971487	26,3307244

Comme pour le domaine spatial, nous avons appliqué une simple substitution LSB aux coefficients DCT quantifiés pour cacher le message secret. Comme nous pouvons le voir dans le tableau 5.17, le nombre de valeurs modifiées par une simple substitution LSB reste significativement plus élevé que le nombre de valeurs modifiées par nos approches. Nous obtenons logiquement un SNR plus élevé pour nos approches. Cependant, étant donné que nous ne dissimulons pas directement dans les valeurs des trames et que les coefficients DCT quantifiés sont manipulés, la capacité de dissimulation de nos approches est bien moindre que celle offerte par la méthode LSB. En effet, notre premier schéma nécessite trois coefficients DCT pour dissimuler deux bits, contrairement à la méthode LSB, qui ne nécessite que deux coefficients (pour dissimuler deux bits).

#### 5.4.3.4 Application de nos approches dans le domaine de la multi-résolution

Afin d'expérimenter nos approches dans le domaine multi-résolution, chaque fichier de notre base de données est décomposé en sous-bandes (AC et DC). Les coefficients AC seront utilisés pour intégrer la marque. Après avoir dissimulé notre message (les informations du patient et les données d'acquisition) les fichiers obtenus ne permettent pas visuellement de détecter des distorsions "Figure 5.12".

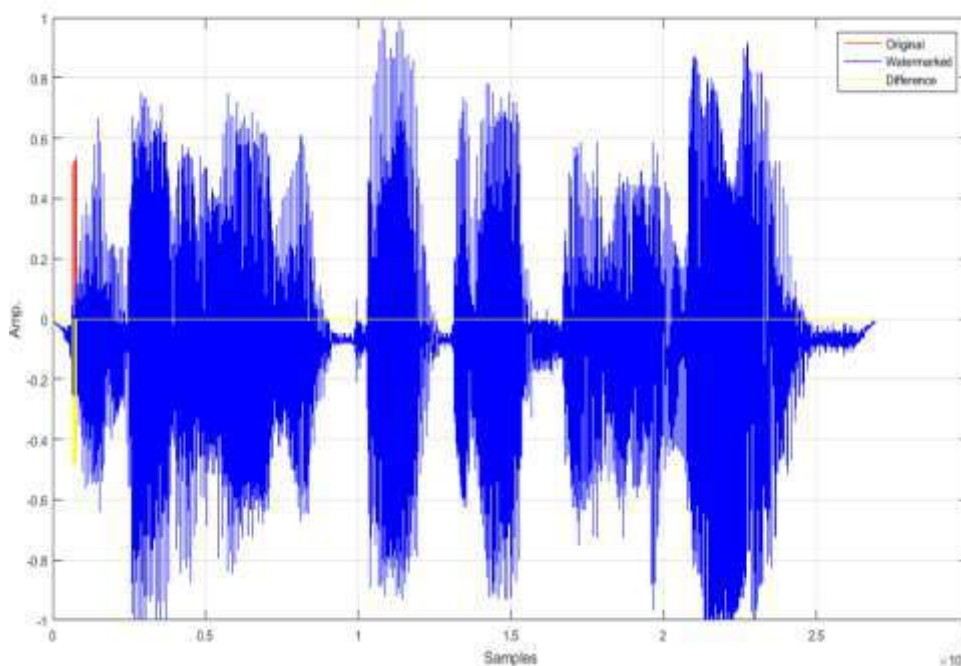


Fig. 5.12 - Différence entre le fichier audio original et le fichier tatoué à l'aide de Sch1.V2

Afin de mesurer les performances et d'évaluer les variantes utilisées, des mesures de similarité et de distorsion entre les fichiers originaux et les fichiers tatoués ont été calculées (tableau 5.18).

Tableau 5.18 - Résultats d'imperceptibilité pour l'insertion dans le domaine multi-résolution

	Sch1.V2	Sch2.V2	Sch2.V1	Sch1.V1	LSB
SNR	42,962	41,918	40,89	40,089	39,718

Afin de comparer nos approches à une simple substitution LSB, nous avons également implémenté et testé une substitution LSB dans les coefficients AC (obtenus par une DWT). Comme pour les résultats obtenus dans les domaines spatial et fréquentiel, une simple substitution LSB génère toujours plus de distorsions que les approches proposées dans le "Tableau 5.18". Ceci est dû au nombre de valeurs modifiées, dans une substitution LSB chaque coefficient a 50% de chance d'être modifié, contrairement à notre approche où un coefficient DWT a 25% de chance d'être modifié, car l'intégration de deux bits se fait en combinant trois coefficients où un seul sera modifié.

### 5.4.3.5 Test d'imperceptibilité

Pour comparer et situer correctement nos approches innovantes par rapport aux travaux récents, pour chaque domaine d'insertion, nous avons objectivement comparé les résultats de distorsion obtenus avec ceux des travaux comparables présentés dans la dernière section du deuxième chapitre de cette thèse, intitulé "*étude de l'état de l'art sur le tatouage audio numérique*".

Tableau 5.19 - Comparaison de l'imperceptibilité avec des travaux connexes

Domaine Spatial	LSB1	Sch1. V1	Sch1. V2	Sch2. V1	Sch2. V2	Hosny & al.[128]	Kundu & Kaur [129]	Devi & Pugazhenthib [130]	
SNR	68,63	69,92	66,92	71,68	72,04	67,63	79,73	73,60	
Domaine de multi-résolution	LSB1	Sch1. V1	Sch1. V2	Sch2. V1	Sch2. V2	Saadi & al.[142]	Kaur & Dutta [143]	Jiang & al.[141]	Hu & al.[144]
SNR	39,72	40,09	42,96	40,89	41,92	41,6001	42,419	40,60	24,63
Domaine Fréquentiel	LSB1	Sch1. V1	Sch1. V2	Sch2. V1	Sch2. V2	Irawati & al.[138]	Jayarani & al.[137]	Karajeh & Maqableh [139]	Xue & al.[140]
SNR	32,39	32,87	32,98	33,19	33,42	26,39	29,15	35,67	24,3813

Le tableau 5.19 montre que le SNR obtenu par nos méthodes innovantes dans les trois domaines d'insertion reste raisonnable par rapport aux résultats de travaux similaires. Dans la première méthode, le processus de dissimulation combine trois informations pour une intégration de deux bits. Sachant qu'une seule valeur peut être modifiée, une de ces trois valeurs est modifiable, ce qui réduit la possibilité d'un éventuel décalage, contrairement à certaines méthodes. Cela signifie normalement une réduction des modifications possibles et donc une réduction de la distorsion du fichier source, ce qui explique correctement le bon taux de SNR obtenu et donc une bonne invisibilité. Cependant, notre première approche nécessite généralement trois éléments possibles pour intégrer deux bits. Dans le domaine spatial, ce problème est parfois insignifiant, puisque tous les éléments du fichier audio sont utilisables pour une bonne insertion. Cependant, la situation se complique dans le domaine de l'insertion fréquentielle, où la quantité de coefficients DCT mesurés est considérablement réduite.

### 5.4.3.6 Test de robustesse

Afin de pouvoir évaluer correctement la robustesse des systèmes proposés (à chaque transformation), différentes attaques potentielles ont été réalisées sur les fichiers tatoués, et les résultats obtenus après la récupération de la marque sont examinés et comparés à des travaux similaires. La corrélation normalisée (NC) est utilisée pour déterminer objectivement la robustesse du modèle de tatouage en comparant la marque originale avec la marque extraite. En général, une  $NC > 0,85$  signifie une forte similarité entre la marque originale et la marque extraite.

Tableau 5.20 - Comparaison de la robustesse avec des travaux connexes

Attaque	Domaine Spatial	Domaine fréquentiel		Domaine Multi-resolution	
	Proposée	Karajeh & Maqableh [139]	Proposée	Kaur & Dutta [143]	Proposée
Bruit gaussien	0.9728	0.9715	0.9830	0.9930	0.9978
Re-échantillonnage	0.9218	0.9619	0.9858	0.9969	0.9963
Filtre passe-bas	0.9504	0.9708	0.9446	0.9429	0.98521
Filtre passe-haut	0.9549	0.9706	0.9686	0.9068	0.9011
Compression MP3	0.9630	0.9720	0.9894	0.9348	0.94325
Recadrage (Cropping)	0.9799	0.9558	0.9575	0.9998	0.9963

Dans le tableau 5.20, nous pouvons clairement affirmer que les scénarios proposés produisent généralement des marques relativement résistantes aux différentes attaques avec une marque de haute qualité. En général, dans le domaine spatial, si les LSB ont été modifiés, il devient impossible d'extraire la marque avec succès. Cependant, dans les systèmes proposés, nous pouvons restaurer instantanément la marque insérée, même après une distorsion des bits LSB. Cette possibilité s'explique par le fait que peu de changements sont effectués efficacement dans les approches proposées, ce qui diminue généralement la possibilité proportionnelle d'un changement significatif. Par conséquent, nous pouvons conclure des résultats obtenus que nos approches innovantes sont suffisamment résistantes à plusieurs attaques malveillantes de marque tout en fournissant une valeur SNR satisfaisante.

### 5.4.4 Analyse et discussion des résultats

Dans cet article, nous avons proposé deux méthodes différentes pour la sécurisation du son des battements de cœur ; ces deux approches ont été utilisées dans trois domaines d'insertion (spatial, fréquentiel et multi-résolution). Dans le domaine spatial, une composition particulière entre les différentes valeurs de données nous a permis de cacher soigneusement deux bits spécifiques en changeant une seule valeur. Le fait remarquable de changer le minimum de données nous a permis de réduire efficacement les distorsions prévisibles, ce qui nous a permis d'obtenir des valeurs SNR acceptables (plus de 66dB). Ces méthodes très simples et peu coûteuses en temps (elles ne nécessitent aucune transformation particulière) sont adaptées au tatouage en temps réel requis dans les environnements en temps réel. Cela permet de protéger les échantillons pendant le transfert (afin de préserver les informations médicales contenues dans le dossier) et d'éviter les confusions entre les échantillons enregistrés puisque l'extraction de la marque garantit que l'échantillon correspond au patient indiqué dans le dossier et évite donc les faux diagnostics.



Nos techniques ont été utilisées dans le domaine fréquentiel, après avoir calculé et quantifié la DCT ; la modification LSB des coefficients quantifiés a été effectuée en comparant la parité entre les éléments consécutifs, permettant ainsi l'insertion des bits de la marque. L'utilisation pratique des transformées rend généralement le message privé plus robuste à une compression efficace, puisqu'elle utilise correctement l'espace similaire exploité pour le codage. En outre, ces schémas élaborés offrent généralement une meilleure résistance aux attaques par filtrage, car les coefficients marqués sont répartis équitablement dans le fichier audio. Les résultats observés sont satisfaisants en comparaison objective avec des travaux similaires où le SNR est supérieur à 33dB en moyenne. Cependant, sur la base de fichiers audio de très petite taille (avec un nombre réduit de coefficients) et puisque notre méthode n'utilise que trois éléments complexes pour intégrer deux bits potentiels, la capacité de nos techniques est considérablement plus faible au niveau des fréquences (environ 4328 à 5770 bits par fichier patient).

Enfin, les méthodes proposées ont été correctement implémentées au domaine multi-résolution en substituant les coefficients complexes des sous-bandes AC locales correctement obtenues. Comme dans les deux domaines précédents, la parité de trois coefficients successifs est typiquement combinée pour intégrer correctement deux bits possibles du message confidentiel. La décomposition fonctionnelle en sous-bandes a typiquement permis d'affiner l'isolement éventuel des composantes de basse fréquence, un espace d'insertion moins sensible tout en préservant le contenu spatial du dossier patient. Les tests d'imperceptibilité révèlent suffisamment des niveaux de SNR adaptés et satisfaisants, soigneusement comparés à certains travaux récents avec un SNR supérieur à 40dB en moyenne pour la première méthode efficace. Contrairement au domaine fréquentiel, le grand nombre de coefficients DWT complexes typiquement générés est plus varié, ce qui augmente l'espace d'intégration de nos approches innovantes.

En comparant rigoureusement les approches que nous avons développées avec les travaux existants dans les trois domaines d'insertion, une conclusion raisonnable peut être tirée : le second modèle présente une imperceptibilité satisfaisante pour l'insertion fréquentielle et spatiale. Cependant, pour les travaux futurs, des améliorations considérables doivent sans aucun doute être apportées pour améliorer de manière significative notre approche flexible de l'insertion dans le domaine multi-résolution en essayant de réduire typiquement les distorsions considérables générées pendant l'insertion. Notre schéma d'insertion dans le domaine spatial devrait également être amélioré pour augmenter progressivement la capacité de dissimulation qui est un critère essentiel pour une méthode de tatouage.

## 5.5 CONCLUSION ET PERSPECTIVES

Dans ce chapitre nous avons présentés nos approches de tatouage numérique aveugle pour protéger des fichiers audio essentiellement du domaine de la télémédecine. Nos approches ont utilisé des techniques de tatouage du domaine spatial, fréquentiel et du domaine de multi-résolution. Chaque approche a été déclinée en plusieurs variantes qui ont été analysées et comparées selon la capacité de dissimulation pour chaque variante et selon l'imperceptibilité en analysant pour chaque méthode les mesures quantitatives de distorsion obtenues. Les expériences réalisées ont montré une robustesse dans l'insertion de la marque et une résistance

aux attaques de filtrage et de compression. Les tests d'imperceptibilité ont donné des niveaux de SNR suffisamment adaptés et satisfaisants, comparés à certains travaux récents. Après analyse des résultats obtenus nous pouvons conclure que nos approches offrent une bonne imperceptibilité et génèrent un échantillon audio tatoué robuste contre diverses attaques avec une marque de haute qualité.

Cependant, pour les travaux futurs, des améliorations considérables doivent sans aucun doute être apportées pour améliorer de manière significative nos approches particulièrement pour l'insertion dans le domaine fréquentiel et le domaine de multi-résolution. L'objectif serait de réduire les distorsions générées lors de l'insertion, et d'augmenter la capacité de dissimulation. Nos schémas d'insertion dans le domaine spatial devraient également être améliorés pour augmenter progressivement la capacité de dissimulation qui est un critère essentiel pour une méthode de tatouage. L'utilisation de fichiers audio de petite taille limite la capacité de nos approches particulièrement dans le domaine fréquentiel.

## 6. Conclusion générale et perspectives

---

### Sommaire

---

6.1 Rappels sur les contributions .....	150
6.2 Perspectives et travaux futurs .....	151

---

---

## 6.1 RAPPELS SUR LES CONTRIBUTIONS

Cette thèse s'inscrit dans la thématique de tatouage audio numérique. L'objectif principal de cette thèse était la conception et l'implémentation des approches de tatouage aveugles qui sont efficaces, robustes et imperceptibles et destinées à protéger des fichiers audio, échangés en télémédecine. Ces fichiers audio contiennent des informations médicales comme les battements de cœurs ou les sons respiratoires. Ce tatouage est réalisé par la dissimulation d'informations secrètes dans le fichier audio hôte et la marque insérée par un tel schéma de tatouage ne doit pas être effaçable, il serait impossible de l'enlever sans avoir recours à la clé ou à la méthode d'insertion. La marque est composée des informations du patient et des données d'acquisition. Ainsi, lors du processus d'extraction, une comparaison entre les informations du patient et la marque indiquera si les données ont été modifiées. Cela permettra de s'assurer que l'échantillon correspond bien au patient indiqué dans l'enregistrement. Trois contributions originales ont été proposées dans cette thèse.

Ainsi, après avoir présenté le contexte d'étude et la problématique (Chapitre1), le son et les fichiers audio numériques et leurs diverses propriétés (Chapitre2), les principaux fondements du tatouage audio numérique, nous avons proposé une classification des différents systèmes de tatouage audio qu'on peut trouver dans la littérature, nous avons montré les techniques d'insertion de marques dans les différents domaines (Chapitre 3), ainsi que les critères et les mesures d'évaluation de la performance des schémas de tatouage audio numérique (Chapitre 4). Nous avons par la suite présenté nos trois contributions (Chapitre 5).

Dans notre première approche intitulée "*Un algorithme de tatouage audio basé sur la transformée de Fourier pour la protection des sons respiratoires*" et afin d'améliorer la sécurité des données échangées en télémédecine, nous avons proposé deux nouveaux schémas de substitution pour le tatouage audio numérique basé sur la transformée de Fourier pour la protection des sons respiratoires. L'insertion de la marque a été réalisée en combinant la parité des valeurs des coefficients successifs, chaque variante représente une combinaison différente. Dans nos expériences, plusieurs variantes de la transformée de Fourier sont utilisées à savoir une Transformée de Fourier Discrète, une Transformée de Fourier Fractionnaire ainsi qu'une Transformée de Fourier Discrète de Quaternion. Pour chaque transformée, les deux variantes de notre schéma de tatouage sont appliquées et les résultats obtenus ont montré que notre approche offre une bonne imperceptibilité et génère un échantillon audio tatoué robuste contre diverses attaques avec une marque de haute qualité. Cependant, l'utilisation de petits fichiers audio pour nos expériences réduit considérablement la capacité de notre approche dans le domaine fréquentiel.

Dans notre deuxième approche intitulée "*Approches spatiales et fréquentielles pour la protection des fichiers audio*" et afin de protéger les fichiers audio, nous avons proposé un nouveau schéma d'insertion pour le tatouage aveugle des fichiers audio. L'objectif est de trouver un compromis entre la capacité et l'imperceptibilité afin de cacher autant de données que possible tout en minimisant la dégradation du fichier. Ce schéma d'insertion est mis en œuvre dans les trois domaines d'insertion : spatial, fréquentiel et multi-résolution. Pour l'insertion dans le domaine spatial, la marque est insérée directement dans les échantillons de données. Pour l'insertion dans le domaine fréquentiel, une transformée en cosinus discrète est appliquée aux trames

---

audio, après l'étape de seuillage et de quantification, la marque est insérée dans les coefficients de la transformée en cosinus discrète pour obtenir le fichier tatoué. Pour l'insertion dans le domaine multi-résolution, une transformée en ondelettes discrètes à un seul niveau est appliquée en utilisant le filtre passe-bas de mise à l'échelle et le filtre passe-haut d'ondelettes. L'insertion de la marque est ensuite effectuée à l'aide des coefficients AC obtenus. Le processus de dissimulation proposé combine trois valeurs pour insérer deux bits et une seule valeur peut être modifiée, ce qui réduit la probabilité de modification contrairement aux autres approches. Cela implique moins de modification et donc moins de distorsion du fichier hôte, ce qui explique le bon rapport signal/bruit obtenu de plus de 59dB pour l'insertion dans le domaine spatial et donc une imperceptibilité raisonnable. Une évaluation de la robustesse du tatouage démontre que les schémas proposés génèrent des échantillons tatoués raisonnablement robustes contre diverses attaques et une marque insérée de haute qualité avec une corrélation croisée normalisée supérieure à 0,9 pour les trois domaines d'insertion.

Dans notre troisième approche intitulée " *Un schéma de tatouage audio basé sur la combinaison de valeur de parité pour la protection des sons de battements de cœur* " et afin d'améliorer la sécurité des données échangées dans le cadre de la télémédecine, nous avons proposé dans ce travail un schéma de tatouage aveugle pour la protection des sons du rythme cardiaque. Deux schémas innovants de tatouage aveugle ont été proposés, chaque schéma élaboré a été décliné en deux variantes spécifiques. L'insertion réussie de la marque se fait typiquement en combinant trois valeurs successives de parité, chaque variante représente une combinaison différente. Ces approches ont été mises en œuvre dans les trois domaines d'insertion : spatial, fréquentiel et multi-résolution. Pour le domaine spatial, la marque est insérée dans les valeurs d'échantillons du fichier. Dans le domaine fréquentiel, les bits de la marque sont substitués au bit le moins significatif du coefficient DCT. Pour l'insertion dans le domaine multi-résolution, après avoir calculé une DWT à un seul niveau, les coefficients de sous-bande AC obtenus sont utilisés pour le processus d'insertion. Après avoir comparé les approches que nous avons utilisées aux divers travaux récents dans les trois domaines utilisés, les résultats obtenus démontrent suffisamment que notre deuxième approche proposée offre une bonne imperceptibilité pour l'insertion observée dans les domaines fréquentiel et spatial. Cependant, l'utilisation de petits fichiers audio pour nos expériences réussies réduit considérablement la capacité de nos méthodes utilisées dans le domaine fréquentiel.

## 6.2 PERSPECTIVES ET TRAVAUX FUTURS

Il y a plusieurs directions pour des recherches futures sur les méthodes proposées introduites dans cette thèse. Dans les travaux futurs, des améliorations considérables doivent sans aucun doute être apportées pour améliorer de manière significative notre approche flexible de l'insertion dans le domaine fréquentiel et le domaine multi-résolution en essayant de réduire typiquement les distorsions considérables générées pendant l'insertion. Notre schéma d'insertion dans le domaine spatial devrait également être amélioré pour augmenter progressivement la capacité de dissimulation qui est un critère essentiel pour une méthode de tatouage.

En plus, un modèle psycho-acoustique pourrait être adopté dans l'utilisation des tests objectifs et perceptifs pour améliorer l'imperceptibilité des méthodes proposées. Nous prévoyons d'utiliser l'algorithme PEAQ (*Perceptual Evaluation of Audio Quality*), pour mesurer la qualité perceptive de nos approches de tatouage. L'algorithme PEAQ renvoie un paramètre appelé ODG (*Objective Difference Grade*) qui se situe entre -4 et 0. Plus la valeur de l'ODG est grande, plus la qualité perceptive est meilleure.

Nous voudrions incorporer des tests de synchronisation et de code de correcteurs d'erreurs pour améliorer la robustesse des méthodes proposées. Certaines attaques modernes telles que l'évanouissement du canal (*channel fading*), le *jitter* et la suppression de paquet seront considérées, car ces attaques sont particulièrement pertinentes dans divers réseaux tels que GSM (Global System for Mobile Communications) et CDMA (Code Division Multiple Access).

La combinaison de plusieurs techniques du domaine spatial et de transformées serait aussi envisageable, pour améliorer la robustesse de nos approches en réduisant au maximum les distorsions générées pendant l'insertion.

Nous pensons à l'ajout d'une couche de chiffrement symétrique ou asymétrique de la marque à insérer avant le processus d'insertion pour renforcer la sécurité des schémas de tatouage et rend la détection de la marque cachée plus difficile. Les données chiffrées présentent généralement un degré élevé d'aléatoire qui s'adapte très bien à la nature des fichiers audio.

Nous pensons également à l'élaboration d'un modèle d'évaluation des schémas de tatouage audio, et pour y arriver nous devrions élargir les tests d'attaques aux attaques de synchronisation et les autres attaques actives et malveillantes comme celles de collision, de suppression de la marque et les attaques cryptographiques.

Enfin, nous compterons élargir notre base de données de test aux fichiers audio un peu plus grands de format WAV et les fichiers audio de nature compressés avec perte comme les fichiers MP3. La complexité de calcul des méthodes proposées sera également réalisée.

## Bibliographie

---

- [1]. I.J. Cox, M.L. Miller, J.A. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, Morgan Kaufmann Publishers. San Francisco, USA, 2008.
- [2] Stefan Katzenbeisser and Fabien A. P Petitcolas, *Information hiding techniques for steganography and digital watermarking*, Published in 2000 in Boston Mass) by Artech house. 2000.
- [3] Neha Singh, Mamta Jain, and Sunil Sharma, "A Survey of Digital Watermarking Techniques," *Int. J. Mod. Commun. Technol. Res. IJMCTR Vol.-1 Issue-6*, Aug. 2013.
- [4] Suraj Kumar Dubey and Vivek Chandra, "Steganography, Cryptography and Watermarking: A Review," *Int. J. Innov. Res. Sci. Eng. Technol. Vol. 6*, pp. 2595–2599, Feb. 2017, [Online]. Available: doi:10.15680/IJIRSET.2017.0602076. 2595
- [5] N.F. Johnson, Z. Duric, and S. Jajodia, *Information Hiding: Steganography and Watermarking - Attacks and Countermeasures*, Kluwer Academic. Boston, USA, 2001.
- [6] F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn, "Information hiding-a survey," *Proc. IEEE*, vol. 87, no. 7, pp. 1062–1078, Jul. 1999, doi: 10.1109/5.771065.
- [7] R.G. van Schyndel, A.Z. Tirkel, and C.F. Osborne, "A digital watermark," in *Proceedings of 1st International Conference on Image Processing*, Nov. 1994, vol. 2, pp. 86–90 vol.2. doi: 10.1109/ICIP.1994.413536.
- [8] G. Widmer, D. Rocchesso, V. Välimäki, C. Erkut, F. Gouyon, D. Pressnitzer, et al, "Sound and music computing: research trends and some key issues.," *J New Music Res* 36, pp. 169–184, 2007.
- [9] Yiqing Lin and Waleed H. Abdulla, "Introduction," in *Audio Watermark A Comprehensive Foundation Using MATLAB*, Springer Cham Heidelberg New York Dordrecht London., 2015.
- [10] L.d.C.T. Gomes, P. Cano, E. Gómez, M. Bonnet, and E. Batlle, "Audio watermarking and fingerprinting: for which applications?," *J New Music Res* 321, pp. 65–81, 2003.
- [11] Saraju P. Mohanty, "Digital Watermarking : A Tutorial Review," *Tech. Rep. Univ. South Fla.*, 1999.
- [12] Yong Xiang, Iynkaran Natgunanathan, Dezhong Peng, Guang Hua, and B. Liu, "Spread Spectrum Audio Watermarking Using Multiple Orthogonal PN Sequences and Variable Embedding Strengths and Polarities," *IEEEACM Trans. Audio Speech Lang. Process. Vol. 26*, pp. 529–539, 2018.
- [13] R. Subhashini and K. BoopathiBagan, "Robust audio watermarking for monitoring and information embedding," presented at the The Fourth International Conference on Signal Processing, Communication and Networking, Chennai, India, 2017.

- 
- [14] I Natgunanathan, Y Xiang, G Hua, G Beliakov, and J Yearwood, "Patchwork-Based Multilayer Audio Watermarking," *IEEEACM Trans. Audio Speech Lang. Process. Vol. 25 Issue 11*, pp. 2176–2187, Nov. 2017.
- [15] Jin-Xia Yang and Dan-Dan Niu, "A novel dual watermarking algorithm for digital audio," presented at the IEEE 17th International Conference on Communication Technology, Chengdu, China, 2017.
- [16] Hongcai-Xu, Xiaobing-Kang, Yajun-Chen, and Yilan-Wang, "Rotation and scale invariant image watermarking based on polar harmonic transforms," *Opt. Vol. 183*, pp. 401–414, Apr. 2019.
- [17] Ratnakirti Roy, Tauheed Ahmed, and Suvamoy Changder, "Watermarking through image geometry change tracking," *Vis. Inform.*, vol. 2, no. 2, pp. 125–135, Jun. 2018, doi: 10.1016/j.visinf.2018.03.001.
- [18] Min-Jae Hwang, JeeSok Lee, MiSuk Lee, and Hong-Goo Kang, "SVD-Based Adaptive QIM Watermarking on Stereo Audio Signals," *IEEE Trans. Multimed. Vol. 20*, pp. 45–54, 2018, [Online]. Available: doi: 10.1109/TMM.2017.2721642
- [19] Wu Weina, "Digital audio blind watermarking algorithm based on audio characteristic and scrambling encryption," presented at the IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference, Chongqing, China, Oct. 2017. [Online]. Available: DOI: 10.1109/IAEAC.2017.8054203
- [20] Zhenghui Liu, Yuankun Huang, and Jiwu Huang, "Patchwork-Based Audio Watermarking Robust Against De-Synchronization and Recapturing Attacks," *IEEE Trans. Inf. Forensics Secur. Vol. 14 Issue 5*, pp. 1171–1180, May 2019, [Online]. Available: DOI: 10.1109/TIFS.2018.2871748
- [21] "Traitement du son," *Wikipédia*. Oct. 25, 2018. Accessed: Dec. 11, 2021. [Online]. Available: [https://fr.wikipedia.org/w/index.php?title=Traitement\\_du\\_son&oldid=153371260](https://fr.wikipedia.org/w/index.php?title=Traitement_du_son&oldid=153371260)
- [22] Grégoire MERCIER, Christian ROUX, and Gilbert MARTINEAU, "Technologies du Multimédia," ENST Bretagne, dpt ITI, BP 832, F-29280 Brest, France., Jan. 15, 2003.
- [23] "Multimédia - Wikipédia." [Online]. Available: <https://fr.wikipedia.org/wiki/Multimédia>
- [24] Brian Moore, "The nature of sound and the structure and function of the auditory system," in *An Introduction to the Psychology of Hearing Sixth Edition*, BRILL., 2013.
- [25] "ANSI S1.1-1994 (R2004) - Acoustical Terminology." <https://webstore.ansi.org/standards/asa/ansis11994r2004> (accessed Nov. 14, 2021).
- [26] Rassol Raissi, "The Theory Behind Mp3," Dec. 2002.
- [27] "Format de fichier audio," *Wikipédia*. Nov. 27, 2021. Accessed: Dec. 17, 2021. [Online]. Available: [https://fr.wikipedia.org/w/index.php?title=Format\\_de\\_fichier\\_audio&oldid=188363668](https://fr.wikipedia.org/w/index.php?title=Format_de_fichier_audio&oldid=188363668)
- [28] "Waveform Audio File Format," *Wikipédia*. Jun. 22, 2021. Accessed: Nov. 05, 2021. [Online]. Available: [https://fr.wikipedia.org/w/index.php?title=Waveform\\_Audio\\_File\\_Format&oldid=184020200](https://fr.wikipedia.org/w/index.php?title=Waveform_Audio_File_Format&oldid=184020200)
- [29] "Audio Interchange File Format," *Wikipédia*. Sep. 13, 2021. Accessed: Nov. 05, 2021. [Online]. Available:
-



- 
- [https://fr.wikipedia.org/w/index.php?title=Audio\\_Interchange\\_File\\_Format&oldid=186301733](https://fr.wikipedia.org/w/index.php?title=Audio_Interchange_File_Format&oldid=186301733)
- [30] “Broadcast Wave Format,” *Wikipédia*. Feb. 25, 2021. Accessed: Nov. 05, 2021. [Online]. Available: [https://fr.wikipedia.org/w/index.php?title=Broadcast\\_Wave\\_Format&oldid=180294560](https://fr.wikipedia.org/w/index.php?title=Broadcast_Wave_Format&oldid=180294560)
- [31] “Core Audio Format,” *Wikipédia*. Mar. 30, 2021. Accessed: Nov. 05, 2021. [Online]. Available: [https://fr.wikipedia.org/w/index.php?title=Core\\_Audio\\_Format&oldid=181410880](https://fr.wikipedia.org/w/index.php?title=Core_Audio_Format&oldid=181410880)
- [32] “Compact Disc Audio track,” *Wikipédia*. Jan. 23, 2020. Accessed: Nov. 05, 2021. [Online]. Available: [https://fr.wikipedia.org/w/index.php?title=Compact\\_Disc\\_Audio\\_track&oldid=166644620](https://fr.wikipedia.org/w/index.php?title=Compact_Disc_Audio_track&oldid=166644620)
- [33] “Sony Global - ATRAC - Overview.” <https://www.sony.net/Products/ATRAC3/overview/> (accessed Nov. 05, 2021).
- [34] “Adaptive Transform Acoustic Coding,” *Wikipédia*. Dec. 31, 2019. Accessed: Nov. 05, 2021. [Online]. Available: [https://fr.wikipedia.org/w/index.php?title=Adaptive\\_Transform\\_Acoustic\\_Coding&oldid=165888682](https://fr.wikipedia.org/w/index.php?title=Adaptive_Transform_Acoustic_Coding&oldid=165888682)
- [35] “FLAC - Free Lossless Audio Codec.” <https://xiph.org/flac/> (accessed Nov. 05, 2021).
- [36] “MP3 (MPEG Layer III Audio Encoding),” May 03, 2017. <https://www.loc.gov/preservation/digital/formats/fdd/fdd000012.shtml> (accessed Nov. 06, 2021).
- [37] “À propos du format audio AC3,” Sep. 14, 2015. <https://www.apowersoft.fr/cest-quoi-un-fichier-ac3.html> (accessed Nov. 06, 2021).
- [38] “MP3,” *Wikipédia*. Oct. 05, 2021. Accessed: Nov. 06, 2021. [Online]. Available: <https://fr.wikipedia.org/w/index.php?title=MP3&oldid=186900274>
- [39] “Mp3PRO,” *Dictionnaires et Encyclopédies sur “Academic.”* <https://fr-academic.com/dic.nsf/frwiki/1196619> (accessed Nov. 06, 2021).
- [40] “Vorbis,” *Wikipédia*. Oct. 11, 2021. Accessed: Nov. 06, 2021. [Online]. Available: <https://fr.wikipedia.org/w/index.php?title=Vorbis&oldid=187059567>
- [41] “Au file format,” *Wikipedia*. Jan. 22, 2021. Accessed: Nov. 06, 2021. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Au\\_file\\_format&oldid=1001922962](https://en.wikipedia.org/w/index.php?title=Au_file_format&oldid=1001922962)
- [42] P. Noll, “MPEG Digital Audio Coding Standards,” 1999. doi: 10.1201/9781420046090-c2.
- [43] P. Noll, “MPEG digital audio coding,” *IEEE Signal Process. Mag.*, vol. 14, no. 5, pp. 59–81, Sep. 1997, doi: 10.1109/79.618009.
- [44] Hwai-Tsu Hu, H.-H. Chou, and Tung-Tsun Lee, “Robust Blind Speech Watermarking via FFT-Based Perceptual Vector Norm Modulation With Frame Self-Synchronization,” *IEEE Access*, vol. 9, pp. 9916–9925, 2021, doi: 10.1109/ACCESS.2021.3049525.
- [45] Jürgen Herre and Sascha Dick, “Psychoacoustic Models for Perceptual Audio Coding—A Tutorial Review,” *Appl. Sci.*, vol. 9, no. 14, Art. no. 14, Jan. 2019, doi: 10.3390/app9142854.
- [46] “ISO/IEC 13818-3:1998(en), Information technology — Generic coding of moving pictures and associated audio information — Part 3: Audio,” Geneva,

- 
- Switzerland 1998. <https://www.iso.org/obp/ui/#iso:std:iso-iec:13818:-3:ed-2:v1:en> (accessed Nov. 09, 2021).
- [47] “ISO/IEC 13818-7:1997, Information technology — Generic coding of moving pictures and associated audio information — Part 7: Advanced Audio Coding (AAC),” *ISO*, Geneva, Switzerland 1997. <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/02/50/25040.html> (accessed Nov. 09, 2021).
- [48] J. Herre and M. Dietz, “MPEG-4 high-efficiency AAC coding [Standards in a Nutshell],” *IEEE Signal Process. Mag.*, vol. 25, no. 3, pp. 137–142, May 2008, doi: 10.1109/MSP.2008.918684.
- [49] Guillaume Fuchs, Werner Oomen, Jeremie Lecomte, Jimmy Lapierre, and Kei Kikuri, “The ISO/MPEG Unified Speech and Audio Coding Standard— Consistent High Quality for All Content Types and at All Bit Rates”, Accessed: Nov. 09, 2021. [Online]. Available: [https://www.academia.edu/20966753/The\\_ISO\\_MPEG\\_Unified\\_Speech\\_and\\_Audio\\_Coding\\_Standard\\_Consistent\\_High\\_Quality\\_for\\_All\\_Content\\_Types\\_and\\_at\\_All\\_Bit\\_Rates](https://www.academia.edu/20966753/The_ISO_MPEG_Unified_Speech_and_Audio_Coding_Standard_Consistent_High_Quality_for_All_Content_Types_and_at_All_Bit_Rates)
- [50] “3GPP. TS 26.441 (V15.0.0): Codec for Enhanced Voice Services (EVS); General overview (3GPP TS 26.441 version 15.0.0 Release 15),” *iTeh Standards Store*, Sophia Antipolis Cedex, France 2018. <https://standards.iteh.ai/catalog/standards/etsi/b891a0a3-170b-420f-bf4c-8f62c1c91812/etsi-ts-126-441-v15-0-0-2018-07> (accessed Nov. 09, 2021).
- [51] Khalid Sayood, *Introduction to Data Compression*. Morgan Kaufmann, 2017.
- [52] “BS.1116 : Methods for the subjective assessment of small impairments in audio systems.” <https://www.itu.int/rec/R-REC-BS.1116-3-201502-I/en> (accessed Nov. 09, 2021).
- [53] “BS.1534 : Method for the subjective assessment of intermediate quality levels of audio systems.” <https://www.itu.int/rec/R-REC-BS.1534> (accessed Nov. 09, 2021).
- [54] Markus Erne, “Perceptual Audio Coders ‘What to listen for,’” presented at the Audio Engineering Society Convention 111, Nov. 2001. Accessed: Nov. 09, 2021. [Online]. Available: <https://www.aes.org/e-lib/browse.cfm?elib=9836>
- [55] Sascha Disch, Steven van de Par, Andreas Niedermeier, Elena Burdiel Pérez, Ane Berasategui Ceberio, and Bernd Edler, “Improved Psychoacoustic Model for Efficient Perceptual Audio Codecs,” presented at the Audio Engineering Society Convention 145, Oct. 2018. Accessed: Nov. 11, 2021. [Online]. Available: <https://www.aes.org/e-lib/online/browse.cfm?elib=19755>
- [56] “ISO/IEC 23003-5:2020, Information technology — MPEG audio technologies — Part 5: Uncompressed audio in MPEG-4 file format,” *ISO*, Jan. 2020. <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/77/77752.html> (accessed Nov. 07, 2021).
- [57] Francis Rumsey, “Virtual Reality, Will It Be a Game-Changer,” *J. Audio Eng. Soc.*, vol. 66, no. 5, pp. 399–402, May 2018, Accessed: Nov. 10, 2021. [Online]. Available: <https://www.aes.org/e-lib/browse.cfm?elib=19569>
- [58] Marek Domański, Olgierd Stankiewicz, Krzysztof Wegner, and Tomasz Grajek, “Immersive visual media — MPEG-I: 360 video, virtual navigation and beyond,” in *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*, May 2017, pp. 1–9. doi: 10.1109/IWSSIP.2017.7965623.
- [59] V.M. Potdar, S. Han, and E.Chang, “A Survey of Digital Image Watermarking Techniques,” presented at the 2005 3rd IEEE International conference on

- 
- Industrial Informatics (INDIN)., Dec. 2005. [Online]. Available: DOI: 10.1109/INDIN.2005.1560462
- [60] Prabhishkek Singh and R S Chadha, "A Survey of Digital Watermarking Techniques, Applications and Attacks," *Int. J. Eng. Innov. Technol. IJEIT Vol. 2 Issue 9*, Mar. 2013.
- [61] N.V. Lalitha, Ch. Srinivasa Rao, and P.V.Y. Jaya Sree, "A REVIEW OF DIGITAL AUDIO WATERMARKING SCHEMES," *J. Crit. Rev. ISSN-2394-5125 Vol 7 Issue 7*, pp. 870–880, 2020, [Online]. Available: DOI: <http://dx.doi.org/10.31838/jcr.07.07.158>
- [62] Ensaf Hussein and Mohamed A. Belal, "Digital Watermarking Techniques, Applications and Attacks Applied to Digital Media: A Survey," *IJERT ISSN 2278-0118 Vol 1 Issue 7*, Sep. 2012.
- [63] Rakesh Ahuja, S S Bedi, and Himanshu Agarwal, "A Survey of Digital Watermarking Scheme," *MIT Int. J. Comput. Sci. Inf. Technol. Vol2 No 1*, p. (52-59), Jan. 2012.
- [64] Alok Chowdhury and Md. Ibrahim Khan, "A Tutorial for Audio Watermarking in the Cepstrum Domain," *Smart Comput. Rev.*, vol. 3, no. 5, Oct. 2013, doi: 10.6029/smarterc.2013.05.003.
- [65] Raniyah Wazirali, Rami Ahmad, Ahmed Al-Amayreh, Mohammad Al-Madi, and Ala' Khalifeh, "Secure Watermarking Schemes and Their Approaches in the IoT Technology: An Overview," *Electron. 2021 10 1744*, 2021, [Online]. Available: <https://doi.org/10.3390/electronics10141744>
- [66] M. Ulutas, G. Ulutas, and V. Nabiyev, "Medical image security and EPR hiding using Shamir's secret sharing scheme," *J Syst Softw*, 2011, doi: 10.1016/J.JSS.2010.11.928.
- [67] Surekha Borra, Rohit Thanki, and Nilanjan Dey, *Digital Image Watermarking Theoretical and Computational Advances*, CRC Press, USA. 2018.
- [68] Guang Hua, Jiwu Huang, Yun Q. Shi, Jonathan Goh, and Vrizlynn L. L. Thing, "Twenty years of digital audio watermarking - A comprehensive review," *Signal Process.*, vol. 128, pp. 222–242, Nov. 2016, doi: 10.1016/j.sigpro.2016.04.005.
- [69] KP Adhiya and SA Patil, "Hiding text in audio using LSB based steganography," *Inf Knowl Manag 238–15*, 2012.
- [70] S.S. Divya and MRM Reddy, "Hiding text in audio using multiple LSB steganography and provide security using cryptography," *Int J Sci Technol Res 1668–70*, Jul. 2012.
- [71] Harish Kumar and Anuradha, "Enhanced LSB technique for audio steganography," in *2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12)*, Jul. 2012, pp. 1–4. doi: 10.1109/ICCCNT.2012.6395978.
- [72] K. Gopalan, "Audio steganography using bit modification," in *2003 International Conference on Multimedia and Expo. ICME '03. Proceedings (Cat. No.03TH8698)*, Jul. 2003, vol. 1, p. I–629. doi: 10.1109/ICME.2003.1220996.
- [73] N.Cvejic and T. Seppanen, "Increasing the capacity of LSB-based audio steganography," *IEEE Workshop Multimed. Signal Process.*, pp. 336–338, Jun. 2003, [Online]. Available: DOI: 10.1109/MMSP.2002.1203314
- [74] N. Cvejic and T. Seppanen, "Increasing Robustness of LSB Audio Steganography Using a Novel Embedding Method," Aug. 2004. [Online]. Available: DOI: 10.1109/ITCC.2004.1286709

- 
- [75] N. Cvejic and T. Seppanen, "Reduced distortion bit-modification for LSB audio steganography," *J. Univers. Comput. Sci. Vol 11 No1*, pp. 56–65, Jun. 2005, [Online]. Available: DOI: 10.1109/ICOSP.2004.1442244
- [76] M.A. Ahmed, M.L.M. Kiah, B.B. Zaidan, and A.A. Zaidan, "A Novel Embedding Method to Increase Capacity and Robustness of Low-bit Encoding Audio Steganography Technique Using Noise Gate Software Logic Algorithm," *J. Appl. Sci. Vol 10*, pp. 59–64, 2010.
- [77] Fatiha Djebbar, Beghdad Ayad, Habib Hamam, and Karim Abed-Meraim, "A view on latest audio steganography techniques," presented at the 2011 International Conference on Innovations in Information Technology, Apr. 2011.
- [78] D. Gruhl, A. Lu, and W. Bender, "Echo hiding," in *Proceeding of Inforomation Hiding Workshop*, 1996, pp. 295–315.
- [79] Y.Erfan and S.Siahpoush, "Robust audio watermarking using improved TS echo hiding," *Digit. Signal Process. Vol 19*, pp. 809–814, Sep. 2009.
- [80] R. Ansari, H. Malik, and A. Khokhar, "Data-hiding in audio using frequency-selective phase alteration," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2004, vol. 5, p. V–389. doi: 10.1109/ICASSP.2004.1327129.
- [81] R. Ansari, H. Malik, and A. Khokhar, "Data-hiding in audio using frequency-selective phase alteration," *IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '04)*, Montreal, Quebec, Canada, pp. 389–392, May 2004.
- [82] H. M. A. Malik, R. Ansari, and A. A. Khokhar, "Robust Data Hiding in Audio Using Allpass Filters," in *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 4, May 2007, pp. 1296–1304.
- [83] David Kahn, "Cryptology and the origins of spread spectrum: Engineers during World War II developed an unbreakable scrambler to guarantee secure communications between Allied leaders; actress Hedy Lamarr played a role in the technology," *IEEE Spectr.*, vol. 21, no. 9, pp. 70–80, Sep. 1984, doi: 10.1109/MSPEC.1984.6370466.
- [84] S. Hernandez-Garay, R. Vazquez-Medina, L. Nino de Rivera, and V. Ponomaryov, "Steganographic communication channel using audio signals," in *12th International Conference on Mathematical Methods in Electromagnetic Theory, (MMET)*, Jul. 2008, pp. 427–429.
- [85] H. Matsuka, "Spread spectrum audio steganography using sub-band phase shifting," Dec. 2006.
- [86] Xin Li and H.H. Yu, "Transparent and robust audio data hiding in subband domain," in *Proceedings International Conference on Information Technology: Coding and Computing (Cat. No.PR00540)*, Mar. 2000, pp. 74–79. doi: 10.1109/ITCC.2000.844186.
- [87] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Syst. J.*, vol. 35, no. 3.4, pp. 313–336, 1996, doi: 10.1147/sj.353.0313.
- [88] In-Kwon Yeo and Hyoung Joong Kim, "Modified Patchwork Algorithm: A Novel Audio Watermarking Scheme," presented at the Proceedings International Conference on Information Technology: Coding and Computing, Aug. 2002. [Online]. Available: DOI: 10.1109/ITCC.2001.918798
- [89] M. Arnold, "Audio watermarking: Features, applications and algorithms," *IEEE Int Conf Multimed. Expo 2000 Vol 2*, pp. 1013–1016, 2000.
- [90] B. Chen and G.W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE*

- 
- Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001, doi: 10.1109/18.923725.
- [91] Hyoung Joong Kim, Yong Hee Choi, Jongwon Seok, and Jinwoo Hong, “Audio Watermarking Techniques,” in *Intelligent Watermarking Techniques*, vol. Volume 7, WORLD SCIENTIFIC, 2004, pp. 185–217. doi: 10.1142/9789812562524\_0008.
- [92] Mohsen Zareian and Hamed Hasani, “Robust image watermarking based on quantization index modulation,” in *2013 IEEE International Conference on Communications (ICC)*, Jun. 2013, pp. 2106–2110. doi: 10.1109/ICC.2013.6654837.
- [93] C.I. Podilchuk and E.J. Delp, “Digital watermarking: algorithms and applications,” *IEEE Signal Process. Mag.*, vol. 18, no. 4, pp. 33–46, Jul. 2001, doi: 10.1109/79.939835.
- [94] IJ Cox, J Kilian, T Leighton, and T Shamoon, “Secure spread spectrum watermarking for images, audio and video.,” in *In: Proceedings of 3rd IEEE international conference on image processing, vol 3. IEEE, Lausanne, Switzerland, Switzerland, 1996*, pp. 243–246,.
- [95] N.Cvejic and T.Seppanen, “Robust audio watermarking in wavelet domain using frequency hopping and patchwork method.,” presented at the In: 3rd International symposium on image and signal processing and analysis, 2003. ISPA 2003. Proceedings of the, vol 1. IEEE, pp 1– 255, Rome, Italy, 2003.
- [96] François Chapeau-Blondeau and Etienne Belin, “Fourier-transform quantum phase estimation with quantum phase noise,” *Signal Process.*, vol. 170, p. 107441, May 2020, doi: 10.1016/j.sigpro.2019.107441.
- [97] Joanna Olbrys and Michal Mursztyn, “Measuring stock market resiliency with Discrete Fourier Transform for high frequency data,” *Phys. Stat. Mech. Its Appl.*, vol. 513, pp. 248–256, Jan. 2019, doi: 10.1016/j.physa.2018.09.028.
- [98] Shuang Zhang, Hua Wang, Jin-gang Gao, and Chun-qi Xing, “Frequency domain point cloud registration based on the Fourier transform,” *J. Vis. Commun. Image Represent.*, vol. 61, pp. 170–177, May 2019, doi: 10.1016/j.jvcir.2019.03.005.
- [99] X.Su, R.Tao, and X.Kang, “Analysis and comparison of discrete fractional fourier transforms,” *Signal Process. Vol. 160*, pp. 284–298, Jul. 2019, [Online]. Available: <https://doi.org/10.1016/j.sigpro.2019.01.019>
- [100] K.Kaur, N.Jindal, and K.Singh, “Riesz fractional order derivative in Fractional Fourier Transform domain: An insight,” *Digit. Signal Process. Vol. 93*, pp. 58–69, Oct. 2019.
- [101] Qingyue Zhang, “Uniqueness guarantees for phase retrieval from discrete windowed fractional Fourier transform,” *Optik*, vol. 158, pp. 1491–1498, Apr. 2018, doi: 10.1016/j.ijleo.2018.01.052.
- [102] Pan Lian, “Uncertainty principle for the quaternion Fourier transform,” *J. Math. Anal. Appl.*, vol. 467, no. 2, pp. 1258–1269, Nov. 2018, doi: 10.1016/j.jmaa.2018.08.002.
- [103] C.Wang, X.Wang, Y.Li, Zhiqiu Xia, and C.Zhang, “Quaternion polar harmonic Fourier moments for color images,” *Inf. Sci. Vol. 450*, pp. 141–156, Jun. 2018.
- [104] S. D. Lin, et al, “Improving the robustness of DCT-based image watermarking against JPEG compression,” *Comput. Stand. Interfaces Vol 32*, pp. 54–60, 2010.
- [105] S. Bedi, et al., “Robust secure SVD based DCT– DWT oriented watermarking technique for image authentication,” in *in International Conference on IT to celebrate S. Charmonman’s 72nd birthday*, 2009, p. 46.1–46.7.

- 
- [106] D. Arya, "A survey of frequency and wavelet domain digital watermarking techniques," *Int. J. Sci. Eng. Res. Vol 1*, 2010.
- [107] N. Ahmed, T. Natarajan, and R. K. Rao, "Discrete Cosine Transform," *IEEE Trans. Comput. Vol 23*, pp. 90–93, 1974.
- [108] I.J. Cox, J. Kilian, F.T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997, doi: 10.1109/83.650120.
- [109] A. B. Watson, "Image Compression Using the Discrete Cosine Transform," *Math. J. Vol 41*, pp. 81–88, 1994.
- [110] Zhao Rui-mei, Lian Hua, Pang Hua-wei, and Hu Bo-ning, "A Watermarking Algorithm by Modifying AC Coefficients in DCT Domain," in *2008 International Symposium on Information Science and Engineering*, Dec. 2008, vol. 2, pp. 159–162. doi: 10.1109/ISISE.2008.20.
- [111] Mahbuba Begum and Mohammad Shorif Uddin, "Digital Image Watermarking Techniques: A Review," *Wwwmdpcomjournalinformationinformation 11 110*, Feb. 2020, [Online]. Available: doi:10.3390/info11020110
- [112] Hai Tao, Li Chongmin, Jasni Mohamad Zain, and Ahmed N. Abdalla, "Robust Image Watermarking Theories and Techniques: A Review," *J. Appl. Res. Technol. Vol12*, Feb. 2014.
- [113] Lee, C., Lee, H., ", "Geometric attack resistant watermarking in wavelet transform domain," in *Optics Express vol. 13, no. 4*, pp. 1307–1321, 2005.
- [114] Wenwu Zhu, Zixiang Xiong, and Ya-Qin Zhang, "Multiresolution watermarking for images and video," *IEEE Trans. Circuits Syst. Video Technol. Vol. 9 Issue 4*, Jun. 1999, [Online]. Available: DOI: 10.1109/76.767121
- [115] Ephraim Feig, "Fast scaled-DCT algorithm," in *Image Processing Algorithms and Techniques*, Jun. 1990, vol. 1244, pp. 2–13. doi: 10.1117/12.19490.
- [116] Dr. Rohit M. Thanki, "Fundamental of Audio Watermarking," in *Advanced Techniques for Audio Watermarking*, Advanced Techniques for Audio Watermarking., 2020. [Online]. Available: DOI: 10.1007/978-3-030-24186-5\_3
- [117] Xin Li and H.H. Yu, "Transparent and robust audio data hiding in cepstrum domain," in *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532)*, Jul. 2000, vol. 1, pp. 397–400 vol.1. doi: 10.1109/ICME.2000.869624.
- [118] K. Gopalan, "Audio steganography by cepstrum modification," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, Mar. 2005, vol. 5, p. v/481-v/484 Vol. 5. doi: 10.1109/ICASSP.2005.1416345.
- [119] Kaliappan Gopalan, "A unified audio and image steganography by spectrum modification," in *2009 IEEE International Conference on Industrial Technology*, Feb. 2009, pp. 1–5. doi: 10.1109/ICIT.2009.4939516.
- [120] K. Gopalan, "A unified audio and image steganography by spectrum modification," *IEEE International Conference on Industrial Technology (ICIT)*, pp. 1–5, Feb. 10, 2009.
- [121] SHI-CHENG LIU and SHINFENG D. LIN, "BCH Code-Based Robust Audio Watermarking in the Cepstrum Domain," *J. Inf. Sci. Eng.* 22 535, pp. 535–543, 2006.
- [122] M. A. Nematollahi, S. A. R. Al-Haddad, and F. Zarafshan, "Blind digital speech watermarking based on Eigen-value quantization in DWT," *J. King Saud*

- 
- Univ. - Comput. Inf. Sci.*, vol. 27, no. 1, pp. 58–67, Jan. 2015, doi: 10.1016/j.jksuci.2014.03.012.
- [123] PK Dhar and T Shimamura, *Advances in audio watermarking based on singular value decomposition*, SpringerBriefs in Electrical and Computer Engineering. 2015. [Online]. Available: DOI 10.1007/978-3-319-14800-7\_5
- [124] Nilanjan Dey, Debalina Biswas, Anamitra Bardhan Roy, Achintya Das, and S. S. Chaudhuri, “DWT-DCT-SVD based blind watermarking technique of gray image in electrooculogram signal,” in *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, Nov. 2012, pp. 680–685. doi: 10.1109/ISDA.2012.6416619.
- [125] V Bhat, I Sengupta, and A Das, “An adaptive audio watermarking based on the singular value decomposition in the wavelet domain,” *Digit. Signal Process* 2061547–1558, 2010.
- [126] Xiang-Yang Wang, Pan-Pan Niu, and Hong-Ying Yang, “A robust digital audio watermarking based on statistics characteristics,” *Pattern Recognit.*, vol. 42, no. 11, pp. 3057–3064, Nov. 2009, doi: 10.1016/j.patcog.2009.01.015.
- [127] X.-Y. Wang and H. Zhao, “A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT,” *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4835–4840, Dec. 2006, doi: 10.1109/TSP.2006.881258.
- [128] A. Hosny, A.Wael, and M.Murtada, “Improving LSB Audio Steganography Using Simulated Annealing for Satellite Telemetry,” presented at the 14th International Computer Engineering Conference, Cairo, Egypt, Egypt, Cairo, Egypt, 2018. [Online]. Available: doi: 10.1109/ICENCO.2018.8636147
- [129] N.Kundu and A.Kaur, “A Secure Approach to Audio Steganography,” *Int. J. Eng. Trends Technol.* 44, Feb. 2017, [Online]. Available: doi:10.35940/ijrte.C4456.098319
- [130] R. Devi and D Pugazhenthib, “Ideal Sampling Rate to reduce distortion in Audio Steganography,” presented at the International Conference on Computational Modeling and Security, Tamil Nadu, India, 2016. [Online]. Available: doi: 10.1016/j.procs.2016.05.185
- [131] Tzuo-Yau Fan, Her-Chang Chao, and Bin-Chang Chieu, “Lossless medical image watermarking method based on significant difference of cellular automata transform coefficient,” *Signal Process Image Commun*, 2019, doi: 10.1016/j.image.2018.09.015.
- [132] S. Haddad, G. Coatrieux, M. Cozic, and D. Bouslimi, “Joint Watermarking and Lossless JPEG-LS Compression for Medical Image Security,” *IRBM*, vol. 38, no. 4, pp. 198–206, Aug. 2017, doi: 10.1016/j.irbm.2017.06.007.
- [133] K.Abdelwahab, M.Saied, W.El-Shafai, , S. El-Rabaie, and F. E. Abd El-Samie, “Efficient SVD-based audio watermarking technique in FRT domain,” *Multimed. Tools Appl.*, Dec. 2019, [Online]. Available: DOI:https://doi.org/10.1007/s11042-019-08023-z
- [134] A.Ortiz, C. Feregrino-Uribe, and J. Garcia-Hernandez, “Framework for audio reversible watermarking robust against content replacement with signal restoration capabilities,” *J. Frankl. Inst. Vol. 356 Issue 12*, pp. 6793–6816, Aug. 2019.
- [135] H.Hu and T.Lee, “High-Performance Self-Synchronous Blind Audio Watermarking in a Unified FFT Framework,” *IEEE Access Vol. 7*, pp. 19063–19076, Jan. 2019.

- 
- [136] M.Seyed, M.Mosleh, and Y. Erfani, "Audio watermarking based on synergy between Lucas regular sequence and Fast Fourier Transform," *Multimed. Tools Appl. Vol. 78*, pp. 22883–22908, Apr. 2019.
- [137] A. Electa Alice Jayarani, Mahabaleswara. R. Bhatt, and D.D Geetha, "Zero Watermarking on Audio Based on STFT," in *2018 International Conference on Computing, Electronics Communications Engineering (iCCECE)*, Aug. 2018, pp. 253–256. doi: 10.1109/iCCECOME.2018.8658846.
- [138] Indrarini Dviah Irawati, Gelar Budiman, and Fuad Ramdhani, "QR-based Watermarking in Audio Sub band Using DCT," presented at the 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), May 2019. [Online]. Available: DOI: 10.1109/ICCEREC.2018.8712108
- [139] H.Karajeh and M.Maqableh, "An imperceptible, robust, and high payload capacity audio watermarking scheme based on the DCT transformation and Schur decomposition," *Analog Integr. Circuits Signal Process. Vol. 99*, pp. 571–583, Jun. 2019, [Online]. Available: doi.org/10.1007/s10470-018-1332-0
- [140] Y. Xue, K.Mu, Y.Wang, Y.Chen, P.Zhong, and J.Wen, "Robust Speech Steganography Using Differential SVD," *IEEE Access Vol. 7*, pp. 153724–153733, Oct. 2019, [Online]. Available: DOI: 10.1109/ACCESS.2019.2948946
- [141] W.Jiang, X.Huang, and Y.Quan, "Audio watermarking algorithm against synchronization attacks using global characteristics and adaptive frame division," *Signal Process. Vol. 162*, pp. 153–160, 2019, [Online]. Available: doi.org/10.1016/j.sigpro.2019.04.017
- [142] S.SAADI, A.MERRAD, and A.BENZIANE, "Novel Secured Scheme for Blind Audio/Speech Norm-Space Watermarking by Arnold Algorithm," *Signal Process. Vol. 154*, pp. 74–86, Jan. 2019, [Online]. Available: doi.org/10.1016/j.sigpro.2018.08.011
- [143] Arashdeep Kaur and Malay Kishore Dutta, "A Blind Watermarking Algorithm for Audio Signals in Multi-Resolution and Singular Value Decomposition," in *2018 4th International Conference on Computational Intelligence Communication Technology (CICT)*, Feb. 2018, pp. 1–5. doi: 10.1109/CICT.2018.8480367.
- [144] Hwai-Tsu Hu, Jieh-Ren Chang, and Shioh-Jyu Lin, "Synchronous blind audio watermarking via shape configuration of sorted LWT coefficient magnitudes," *Signal Process.*, vol. 147, pp. 190–202, Jun. 2018, doi: 10.1016/j.sigpro.2018.02.001.
- [145] Xing He, "Introduction of Human Auditory System and Psychoacoustics," in *Signal Processing, Perceptual Coding and Watermarking of Digital Audio: Advanced Technologies and Models*, Information Science Reference (an imprint of IGI Global)., USA: SRS Labs Inc., USA, 2012.
- [146] Ingemar Cox, Matthew Miller, Jeffrey Bloom, and Mathew Miller, *Digital Watermarking*, CA: Academic Press. San Francisco, 2002.
- [147] T. Painter and A. Spanias, "Perceptual coding of digital audio," in *Proceedings of the IEEE*, 88(4), Apr. 2000, pp. 451–513. [Online]. Available: doi:10.1109/5.842996
- [148] E.Zwicker and H.Fastl, *Psychoacoustics: Facts and Models*, New York, NY: Springer-Verlag. 1990.
- [149] Brian Moore, *An Introduction to the Psychology of Hearing*, Brill. 2013.



- 
- [150] E. Zwicker and E. Terhardt, "Analytical expressions for critical- band rate and critical bandwidth as a function of frequency," *J. Acoust. Soc. Am.*, vol. 68, no. 5, pp. 1523–1525, Nov. 1980, doi: 10.1121/1.385079.
- [151] B. C. J. Moore and B. R. Glasberg, "A Revision of Zwicker's Loudness Model," *Acta Acust. United Acust.*, vol. 82, no. 2, pp. 335–345, Mar. 1996.
- [152] B. C. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.*, vol. 74, no. 3, pp. 750–753, Sep. 1983, doi: 10.1121/1.389861.
- [153] A.Spanias, T.Painter, V.Atti, and J.V.Candy, "Audio Signal Processing and Coding," *J. Acoust. Soc. Am.* 1221, Jul. 2007, [Online]. Available: DOI: 10.1121/1.2737746
- [154] Manfred R. Schroeder and Bishnu S. Atal, "Stochastic coding of speech signals at very low bit rates," in *In Proceedings of the IEEE International Conference on Communications*, Amsterdam, Netherlands., 1984, pp. 1610–1613.
- [155] H. Fastl and E. Zwicker, "Critical Bands and Excitation," in *Psychoacoustics: Facts and Models*, H. Fastl and E. Zwicker, Eds. Berlin, Heidelberg: Springer, 2007, pp. 149–173. doi: 10.1007/978-3-540-68888-4\_6.
- [156] E.Zwicker and U.Zwicker, "Audio engineering and psychoacoustics: matching signals to the final receiver, the human auditory system," *J. Audio Eng. Soc. Audio Eng. Soc.* 39, pp. 115-126., Published 1991, [Online]. Available: DOI:10.1016/B978-155860651-7/50088-7Corpus ID: 60476305
- [157] RhonaP. Hellman, "Asymmetry of masking between noise and tone," *Percept. Psychophys. Vol. 11*, pp. 241–246, 1972.
- [158] B.Scharf, "Critical Bands," in *Foundations of Modern Auditory Theory*, Editor : Jerry Tobias, New York, NY: Academic Press., USA, 1970, pp. 157–202.
- [159] Ali N.Akansu and Mark J.T.Smith, *Subband and wavelet transforms, design and applications.*, Norwell, MA: Kluwer Academic Publishers. 1996.
- [160] Xiangui Kang, Rui Yang, and Jiwu Huang, "Geometric Invariant Audio Watermarking Based on an LCM Feature," *IEEE Trans. Multimed.*, vol. 13, no. 2, pp. 181–190, Apr. 2011, doi: 10.1109/TMM.2010.2098850.
- [161] Y Xiang, I Natgunanathan, S Guo, W Zhou, and S Nahavandi, "Patchwork-based audio watermarking method robust to de-synchronization attacks," *IEEEACM Trans Audio Speech Lang Process* 229, pp. 1413–1423, Sep. 2014, [Online]. Available: DOI: 10.1109/TASLP.2014.2328175
- [162] M. Steinebach *et al.*, "StirMark Benchmark: Audio watermarking attacks," presented at the In: Int. Conference on Information Technology: Coding and Computing (ITCC 2001), April 2 - 4, Las Vegas, Nevada, S. 49 - 54, ISBN 0-7695-1062-0,2001, Aug. 2002.
- [163] A. Zaidi, R. Boyer, and P. Duhamel, "Audio watermarking under desynchronization and additive noise attacks," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 570–584, Feb. 2006, doi: 10.1109/TSP.2005.861106.
- [164] A. Robert and J. Picard, "On the use of masking models for image and audio watermarking," *IEEE Trans. Multimed.*, vol. 7, no. 4, pp. 727–739, Aug. 2005, doi: 10.1109/TMM.2005.846781.
- [165] G. Doerr, J.-L. Dugelay, and D. Kirovski, "On the Need for Signal-Coherent Watermarks," *IEEE Trans. Multimed.*, vol. 8, no. 5, pp. 896–904, Oct. 2006, doi: 10.1109/TMM.2006.879917.

- 
- [166] D. Kirovski, F. A.P. Petitcolas, and Z. Landau, “The Replacement Attack,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 6, pp. 1922–1931, Aug. 2007, doi: 10.1109/TASL.2007.900088.
- [167] D. Kirovski and F.A.P. Petitcolas, “Blind pattern matching attack on watermarking systems,” *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1045–1053, Apr. 2003, doi: 10.1109/TSP.2003.809383.
- [168] Yiqing Lin and Waleed H. Abdulla, “Audio Watermarking Techniques,” in *Audio Watermark A Comprehensive Foundation Using MATLAB*, Springer Cham Heidelberg New York Dordrecht London., 2015. [Online]. Available: DOI 10.1007/978-3-319-07974-5
- [169] P. Liew and M. Armand, “Inaudible watermarking via phase manipulation of random frequencies,” *Multimed. Tools Appl* 353 357–377, 2007.
- [170] Yiqing Lin and Waleed H. Abdulla, “Appendix F : Basic Robustness Test,” in *Audio Watermark A Comprehensive Foundation Using MATLAB*, Springer Cham Heidelberg New York Dordrecht London., 2015.
- [171] F. RAYNAL, F.A.P. PETITCOLAS, and C. FONTAINE, “Évaluation automatique des méthodes de tatouage,” *Trait. Signal — Vol. 18 — N 4 — Spéc. 2001*, 2001.
- [172] J.C. Vorbruggen and F. Cayre, “The Certimark benchmark: architecture and future perspectives,” in *Proceedings. IEEE International Conference on Multimedia and Expo*, Aug. 2002, vol. 2, pp. 485–488 vol.2. doi: 10.1109/ICME.2002.1035651.
- [173] Shelby Pereira, Sviatoslav Voloshynovskiy, Maribel Madueno, Stéphan Marchand-Maillet, and Thierry Pun, “Second Generation Benchmarking and Application Oriented Evaluation,” in *Information Hiding*, Berlin, Heidelberg, 2001, pp. 340–353. doi: 10.1007/3-540-45496-9\_25.
- [174] Peter Meerwald and Shelby Pereira, “Attacks, applications, and evaluation of known watermarking algorithms with Checkmark,” in *Security and Watermarking of Multimedia Contents IV*, Apr. 2002, vol. 4675, pp. 293–304. doi: 10.1117/12.465287.
- [175] “SDMI Portable Device Specification, Part 1 (Version 1.0). SDMI (1999) [Online]. <http://ntrg.cs.tcd.ie/undergrad/4ba2.01/group10/technology.html>,” *The blue social bookmark and publication sharing system- BibSonomy Genealogy*, 1999.
- [176] “Call for Proposals for Phase II Screening Technology (Version 1.0). SDMI (2000) [Online]. <http://ntrg.cs.tcd.ie/undergrad/4ba2.01/group10/technology.html>.” <http://link.springer.com> › bbm:978-3-319-07974-5 › 1.pdf, 2000.
- [177] “Announcement of Evaluation Test Results for “STEP 2000”. JASRAC and NRI (2000) [Online], <http://www.jasrac.or.jp/watermark/ehoukoku.htm>.” <http://link.springer.com> › bbm:978-3-319-07974-5 › 1.pdf, 2000.
- [178] Yong Xiang, Guang Hua, and Bin Yan, “Human Auditory System and Perceptual Quality Measurement,” in *Digital Audio Watermarking Fundamentals, Techniques and Challenges*, SpringerBriefs in Electrical and Computer Engineering., 2017.
- [179] S. A. Gelfand, *Hearing : an introduction to psychological and physiological acoustics*, Acoustics, 3rd Edition, Marcel Dekker, Basel, Switzerland. 1998.
- [180] M. Unoki, Kuniaki Imabeppu, Daiki Hamada, Atsushi Haniu, and Ryota Miyauchi, “Embedding Limitations with Digital-audio Watermarking Method

- 
- Based on Cochlear Delay Characteristics,” *J Inf Hiding Multim Signal Process*, 2011.
- [181] Y.Xian, D.Peng, I.Natgunanathan, and W.Zhou, “Effective pseudonoise sequence and decoding function for imperceptibility and robustness enhancement in time-spread echo-based audio watermarking,” *IEEE Trans Multimed.* 131, pp. 2–13, 2011.
- [182] Byeong-Seob Ko, R. Nishimura, and Y. Suzuki, “Time-spread echo method for digital audio watermarking,” *IEEE Trans. Multimed.*, vol. 7, no. 2, pp. 212–221, Apr. 2005, doi: 10.1109/TMM.2005.843366.
- [183] Zhe-Ming Lu, Bin Yan, and Sheng-He Sun, “Watermarking Combined with CELP Speech Coding for Authentication,” *IEICE Trans. Inf. Syst.*, vol. E88-D, no. 2, pp. 330–334, Feb. 2005, Accessed: Nov. 03, 2021. [Online]. Available: [https://search.ieice.org/bin/summary.php?id=e88-d\\_2\\_330&category=D&year=2005&lang=E&abst=d\\_2\\_330](https://search.ieice.org/bin/summary.php?id=e88-d_2_330&category=D&year=2005&lang=E&abst=d_2_330)
- [184] Jon Boley and Michael Lester, “Statistical Analysis of ABX Results Using Signal Detection Theory,” presented at the Audio Engineering Society Convention 127, Oct. 2009. Accessed: Nov. 04, 2021. [Online]. Available: <https://www.aes.org/e-lib/online/browse.cfm?elib=15022>
- [185] WB Kleijn and KK Paliwal, *Speech coding and synthesis*, Elsevier Science Inc., New York. 1995.
- [186] ITU-T, “Recommendation p.800: Methods for objective and subjective assessment of quality. <http://www.itu.int/>,” 1996. [Online]. Available: <http://www.itu.int/>
- [187] Sheetal Shelke, Mangal Patil, and J. S. Chitode, “DWT-FFT Based Audio Watermarking Algorithm for Copyright,” *Int. J. Eng. Adv. Technol. IJEAT ISSN 2249 – 8958 Vol.-5 Issue-2*, Dec. 2015.
- [188] Yi Hu and Philipos C. Loizou, “Evaluation of Objective Quality Measures for Speech Enhancement,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008, doi: 10.1109/TASL.2007.911054.
- [189] ITU-T, “Recommendation ITU-R BS. 1387-1: Method for objective measurements of perceived audio quality,” 2001 1998. [Online]. Available: <http://www.itu.int/>
- [190] R Huber and B Kollmeier, “PEMO-Q: A new method for objective audio quality assessment using a model of auditory perception,” *IEEE Trans Audio Speech Lang Process* 146, pp. 1902–1911, Oct. 2006, [Online]. Available: DOI: 10.1109/TASL.2006.883259
- [191] WC Treurniet and GA Soulodre, “Evaluation of the ITU-R objective audio quality measurement method,” *JAES Vol. 48 Issue 3*, pp. 164–173, Mar. 2000.
- [192] NK Kalantari, MA Akhaee, SM Ahadi, and H Amindavar, “Robust multiplicative patchwork method for audio watermarking,” *IEEE Trans Audio Speech Lang Process* 176, p. :1133-1141, 2009.
- [193] Guang Hua, Jonathan Goh, and Vrizlynn. L. L. Thing, “Time-Spread Echo-Based Audio Watermarking With Optimized Imperceptibility and Robustness,” *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 2, pp. 227–239, Feb. 2015, doi: 10.1109/TASLP.2014.2387385.
- [194] C. Baras, N. Moreau, and P. Dymarski, “Controlling the inaudibility and maximizing the robustness in an audio annotation watermarking system,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 5, pp. 1772–1782, Sep. 2006, doi: 10.1109/TASL.2006.879808.
-

- 
- [195] XY Wang, PP Niu, and HY Yang, "A robust, digital-audio watermarking method," *IEEE Multimed. Vol. 16 Issue 3*, pp. 60–69, Sep. 2009.
- [196] XY Wang, W Qi, and PP Niu, "A new adaptive digital audio watermarking based on support vector regression," *IEEE Trans Audio Speech Lang Process* 158, pp. 2270–2277, Oct. 2007.
- [197] Oscar T.-C. Chen and Wen-Chih Wu, "Highly Robust, Secure, and Perceptual-Quality Echo Hiding Scheme," *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 3, pp. 629–638, Mar. 2008, doi: 10.1109/TASL.2007.913022.
- [198] F.A.P. Petitcolas, "Watermarking schemes evaluation," *IEEE Signal Process. Mag.*, vol. 17, no. 5, pp. 58–64, Sep. 2000, doi: 10.1109/79.879339.
- [199] Alok Kumar Chowdhury and Md. Ibrahim Khan, "A Tutorial for Audio Watermarking in the Cepstrum Domain," *Smart Computing Review*, vol. 3, no., Oct. 05, 2013. [Online]. Available: DOI: 10.6029/smarterc.2013.05.003
- [200] NN Hurrah, S.A Parah, J.A. Sheikh, F. Al-Turjman, and K. Muhammad, "Secure data transmission framework for confidentiality in IoTs," *Ad Hoc Netw* 95 101989, 2019, [Online]. Available: <https://doi.org/10.1016/j.adhoc.2019.101989>
- [201] Ali Fatahbeygi and Fardin Akhlaghian Tab, "A highly robust and secure image watermarking based on classification and visual cryptography," *J. Inf. Secur. Appl.*, vol. 45, pp. 71–78, Apr. 2019, doi: 10.1016/j.jisa.2019.01.005.
- [202] O.Kwon, S.Choi, and B.Lee, "A Watermark-Based Scheme for Authenticating JPEG Image Integrity," *IEEE Access Vol. 6*, pp. 46194–46205, 2018.
- [203] Y. Zhang and Y.Sun, "An image watermarking method based on visual saliency and contourlet transform," *Opt. Vol. 186*, pp. 379–389, Jun. 2019.
- [204] T.Araghi and A.Manaf, "An enhanced hybrid image watermarking scheme for security of medical and non-medical images based on DWT and 2-D SVD," *Future Gener. Comput. Syst. Vol. 101*, pp. 1223–1246, Dec. 2019.
- [205] B.Xiao, J.Luo, X.Bi, W.Li, and B.Chen, "Fractional discrete Tchebyshev moments and their applications in image encryption and watermarking," *Inf. Sci. Vol. 516*, pp. 545–559, Apr. 2020.
- [206] X.Liu *et al.*, "A Novel Robust Reversible Watermarking Scheme for Protecting Authenticity and Integrity of Medical Images," *IEEE Access Vol. 7*, pp. 76580–76598, 2019.
- [207] K.Pexas, G. Karybali, and E.Kalligeros, "Optimization and Hardware Implementation of Image and Video Watermarking for Low-Cost Applications," *IEEE Trans. Circuits Syst. Vol. 66*, pp. 2088–2101, 2019.
- [208] Bruno M. Rocha *et al.*, "An open access database for the evaluation of respiratory sound classification algorithms," *Physiol. Meas.*, vol. 40, no. 3, p. 035001, Mar. 2019, doi: 10.1088/1361-6579/ab03ea.
- [209] D.Renza, M. Ballesteros, and C.Lemus, "Authenticity verification of audio signals based on fragile watermarking for audio forensics," *Expert Syst. Appl. Vol. 91*, pp. 211–222, Jan. 2018.
- [210] M.Mulimani and G. Koolagudi, "Extraction of MapReduce-based features from spectrograms for audio-based surveillance," *Digit. Signal Process. Vol. 87*, pp. 1–9, Apr. 2019.
- [211] A. Khaldi, "Steganographic Techniques Classification According to Image Format," *Int. Ann. Sci. Vol. 8*, pp. 143–149, 2019, [Online]. Available: doi: 10.21467/ias.8.1.143-149

- 
- [212] Q.Su and B.Chen, "A novel blind color image watermarking using upper Hessenberg matrix," *AEU - Int. J. Electron. Commun. Vol. 78*, pp. 64–71, Aug. 2017, [Online]. Available: doi.org/10.1142/S0219467809003514
- [213] Y. Yao, W. Zhang, H. Wang, H. Zhou, N. Yu, "Content-adaptive reversible visible watermarking in encrypted images," *Signal Process. Vol. 164*, pp. 386–401, Nov. 2019, [Online]. Available: doi.org/10.1016/j.sigpro.2019.06.034
- [214] Y. Gangadhar, V. S. Giridhar Akula, and P. Chenna Reddy, "An evolutionary programming approach for securing medical images using watermarking scheme in invariant discrete wavelet transformation," *Biomed. Signal Process. Control Vol. 43*, pp. 31–40, May 2018, [Online]. Available: DOI: 10.1016/j.bspc.2018.02.007
- [215] "BBC Sound Archive," *Wikipedia*. May 12, 2021. Accessed: Nov. 29, 2021. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=BBC\\_Sound\\_Archive&oldid=1022853716](https://en.wikipedia.org/w/index.php?title=BBC_Sound_Archive&oldid=1022853716)
- [216] Mohammed Mahdi Hashim, Mohd Shafry Mohd Rahim, Fadil Abass Johi, Mustafa Sabah Taha, and Hassan Salman Hamad, "Performance evaluation measurement of image steganography techniques with analysis of LSB based on variation image formats," *Int. J. Eng. Technol. Vol. 7*, pp. 3505–3514, Dec. 2018, [Online]. Available: doi: 10.14419/ijet.v7i4.17294
- [217] A. Khaldi, "Diffie-Hellman Key Exchange through Steganographed Images," *Rev. Direito Estado E Telecomunicações Vol. 10*, pp. 147–160, May 2018, [Online]. Available: doi.org/10.26512/lstr.v10i1.21504
- [218] F.Kahlessenane, A.Khaldi, and S.Euschi, "A robust blind color image watermarking based on Fourier transform domain," *Opt. Vol. 208*, Apr. 2020, [Online]. Available: doi.org/10.1016/j.ijleo.2020.164562
- [219] A. Khaldi, "A lossless blind image data hiding scheme for semi-fragile image watermark," *Int. J. Comput. Vis. Robot.*, 2020, [Online]. Available: doi: 10.1504/IJCVR.2020.10029218
- [220] Ed King, "Heartbeat Sounds," *Dataset - Heartbeat Sounds, Stanford University, Stanford, California, United States*. <https://www.kaggle.com/kinguistics/heartbeat-sounds?>
- [221] Euschi Salah, Khaldi Amine, Kafi Redouane, and Kahlessenane Fares, "A Fourier transform based audio watermarking algorithm," *Appl. Acoust.*, vol. 172, p. 107652, Jan. 2021, doi: 10.1016/j.apacoust.2020.107652.
- [222] B. Rocha *et al.*, "A Respiratory Sound Database for the Development of Automated Classification," 2017. doi: 10.1007/978-981-10-7419-6\_6.
- [223] "ICBHI 2017 Challenge." [https://bhichallenge.med.auth.gr/ICBHI\\_2017\\_Challenge](https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge) (accessed Dec. 11, 2021).
- [224] "BBC Sound Effects." <https://sound-effects.bbcrewind.co.uk/> (accessed Dec. 29, 2021).
- [225] Peter Bentley, Glenn Nordehn, Miguel Coimbra, Shie Mannor, and Rita Getz, "Classifying Heart Sounds Challenge." <http://www.peterjbentley.com/heartchallenge/index.html> (accessed Dec. 10, 2021).

## Nos publications

---

### PUBLICATIONS INTERNATIONALES

1. Euschi Salah, Khaldi Amine, Kafi Redouane, and Kahlessenane Fares, “A Fourier transform based audio watermarking algorithm,” *Applied Acoustics*, vol. 172, p. 107652, Jan. 2021, doi: 10.1016/j.apacoust.2020.107652.
2. Euschi Salah, Khaldi Amine, Kafi Med Redouane, and Kahlessenane Fares, “Spatial and Frequency Approaches for Audio File Protection,” *J CIRCUIT SYST COMP*, vol. 30, no. 12, p. 2150210, Sep. 2021, doi: 10.1142/S0218126621502108.
3. Euschi Salah, Khaldi Amine, Kafi Redouane, and Kahlessenane Fares, “A Value Parity Combination-Based Scheme for Heartbeat Sounds Protection,” in *Internet of Things*, CRC Press, 2021.

### CONFERENCES NATIONALES

1. Euschi Salah, Khaldi Amine and Kafi Redouane, “A blind frequency based scheme for audio sounds watermarking“, *Conférence Nationale en Intelligence Artificielle et Technologie de l’Information CNIATI’20* , 24 may 2021.

### PUBLICATIONS INTERNATIONALES COMME CO-AUTEUR

1. Fares Kahlessenane, Amine Khaldi, Med Redouane Kafi, and Salah Euschi, “A color value differentiation scheme for blind digital image watermarking,” *Multimed Tools Appl*, vol. 80, no. 13, pp. 19827–19844, May 2021, doi: 10.1007/s11042-021-10713-6.
2. F.Kahlessenane, A.Khaldi, and S.Euschi, “A robust blind color image watermarking based on Fourier transform domain,” *Optik*, Volume 208, Apr. 2020, [Online]. Available: doi.org/10.1016/j.ijleo.2020.164562
3. Fares Kahlessenane, Amine Khaldi, Redouane Kafi, and Salah Euschi, “A robust blind medical image watermarking approach for telemedicine applications,” *Cluster Comput*, vol. 24, no. 3, pp. 2069–2082, Sep. 2021, doi: 10.1007/s10586-020-03215-x.
4. Fares Kahlessenane, Amine Khaldi, Med Redouane Kafi, Narima Zermi, and Salah Euschi, “A value parity combination based scheme for retinal images watermarking,” *Opt Quant Electron*, vol. 53, no. 3, p. 161, Mar. 2021, doi: 10.1007/s11082-021-02793-3.

- 
5. Kahlessenane Fares, Amine Khaldi, Kafi Redouane, and Euschi Salah, "DCT & DWT based watermarking scheme for medical information security," *Biomedical Signal Processing and Control*, vol. 66, p. 102403, Apr. 2021, doi: 10.1016/j.bspc.2020.102403.
  6. Kahlessenane Fares, Amine Khaldi, Kafi Redouane, and Euschi Salah, "A DWT based watermarking approach for medical image protection," *Journal of Ambient Intelligence and Humanized Computing* volume 12, pages2931–2938 (2021), DOI: 10.1007/s12652-020-02450-9.
  7. Narima Zermi, Amine Khaldi, Med Redouane Kafi, Fares Kahlessenane, and Salah Euschi, "A lossless DWT-SVD domain watermarking for medical information security," *Multimed Tools Appl*, vol. 80, no. 16, pp. 24823–24841, Jul. 2021, doi: 10.1007/s11042-021-10712-7.
  8. Narima Zermi, Amine Khaldi, Mohamed Redouane Kafi, Fares Kahlessenane, and Salah Euschi, "An SVD Values Ordering Scheme for Medical Image Watermarking," *Cybernetics and Systems*, vol. 0, no. 0, pp. 1–16, Oct. 2021, doi: 10.1080/01969722.2021.1983700.
  9. Narima Zermi, Amine Khaldi, Med Redouane Kafi, Fares Kahlessenane, and Salah Euschi, "Robust SVD-based schemes for medical image watermarking," *Microprocessors and Microsystems*, vol. 84, p. 104134, Jul. 2021, doi: 10.1016/j.micpro.2021.104134.
  10. Narima Zermi, Amine Khaldi, Med Redouane Kafi, Fares Kahlessenane, and Salah Euschi, "A DWT-SVD based robust digital watermarking for medical image security," *Forensic Science International* 320(January):110691, January 2021, DOI: 10.1016/j.forsciint.2021.110691.
  11. Khaldi Amine, Kahlessenane Fares, Kafi Med Redouane, and Euschi Salah, "Medical Image Watermarking for Telemedicine Application Security," *J CIRCUIT SYST COMP*, p. 2250097, Oct. 2021, doi: 10.1142/S0218126622500979.

## Table des annexes

---

### Table des annexes

1	<b>Annexe 1 : Base de données des sons respiratoires .....</b>	171
2	<b>Annexe 2 : Base de données des Effets Sonores de la BBC .....</b>	174
3	<b>Annexe 3 : Base de données des sons de battements cardiaques .....</b>	176



## **Annexe 1 – Base de données des sons respiratoires**

---

### **CONTEXTE**

L'analyse automatique des bruits respiratoires a suscité un grand intérêt de la part des chercheurs au cours des dernières décennies. La classification automatique des bruits respiratoires a le potentiel de détecter des anomalies dans les premiers stades d'un dysfonctionnement respiratoire et donc d'améliorer l'efficacité de la prise de décision [208]. Cependant, l'existence d'une grande base de données publiquement disponible, dans laquelle de nouveaux algorithmes peuvent être mis en œuvre, évalués et comparés, fait toujours défaut et est essentielle pour les développements futurs dans ce domaine.

Dans le cadre de la Conférence internationale sur l'informatique biomédicale et de santé (ICBHI), le premier défi scientifique a été organisé avec pour objectif principal de développer des algorithmes capables de caractériser des enregistrements de sons respiratoires provenant d'environnements cliniques et non cliniques. La base de données a été créée par deux équipes de recherche au Portugal et en Grèce, et elle comprend 920 enregistrements acquis auprès de 126 sujets et deux ensembles d'annotations. L'un des ensembles contient 6898 cycles respiratoires annotés, certains comprenant des crépitements, des sifflements ou une combinaison des deux, et d'autres ne comportant aucun bruit respiratoire fortuit. Dans l'autre ensemble, les emplacements précis de 10 775 événements de crépitements et de sifflements ont été annotés. Les cycles ont été annotés par des experts. Les enregistrements ont été recueillis à l'aide d'un équipement hétérogène et leur durée variait de 10 à 90 s. Les emplacements du thorax à partir desquels les enregistrements ont été acquis ont également été fournis. Les niveaux de bruit dans certains cycles respiratoires étaient élevés, ce qui simulait les conditions réelles et rendait le processus de classification plus difficile [222]. La création et la diffusion publique de cette base de données seront utiles à la communauté des chercheurs et pourraient attirer l'attention sur le problème de la classification des sons respiratoires [208].

La base de données des sons respiratoires a été initialement compilée pour soutenir le défi scientifique organisé lors de l'Int. Conf. on Biomedical Health Informatics - ICBHI 2017. La version actuelle de cette base de données est mise gratuitement à la disposition de la recherche et contient à la fois le jeu de données public et le jeu de données privé du défi ICBHI.

## CONTENU

Chaque nom de fichier est divisé en 5 éléments, séparés par des caractères de soulignement (  ) [223].

1. Numéro du patient (**101,102,....,226**)
2. Index de l'enregistrement
3. Emplacement du thorax
  - a. Trachée (**Tc**)
  - b. Antérieur gauche (**Al**)
  - c. Antérieur droit (**Ar**)
  - d. Postérieur gauche (**Pl**)
  - e. Postérieure droite (**Pr**)
  - f. Latéral gauche (**Ll**)
  - g. Latéral droit (**Lr**)
4. Mode d'acquisition
  - a. séquentiel/unique (**sc**),
  - b. simultané/multi-canal (**mc**)
5. Equipement d'enregistrement
  - a. Microphone AKG C417L (**AKGC417L**)
  - b. Stéthoscope Littmann Classic II SE de 3M (**LittC2SE**),
  - c. Stéthoscope électronique 3M Litmmann 3200 (**Litt3200**),
  - d. Stéthoscope électronique Meditron Master Elite de WelchAllyn (**Meditron**)

Les fichiers d'annotation comprennent quatre colonnes :

- Début du (des) cycle(s) respiratoire(s)
- Fin de cycle(s) respiratoire(s)
- Présence/absence de crépitements (présence=1, absence=0)
- Présence/absence de sifflements (présence=1, absence=0)

Les abréviations utilisées dans le dossier de diagnostic sont les suivantes :

- COPD: Chronic Obstructive Pulmonary Disease
- LRTI: Lower Respiratory Tract Infection

- URTI: Upper Respiratory Tract Infection

## **TELECHARGEMENT**

Cette base de données est disponible gratuitement pour la recherche et peut être téléchargée au format zip à partir de [ce lien](#) ou vous pouvez parcourir les fichiers de la base de données sur [cette page](#). En outre, vous pouvez télécharger les fichiers contenant les informations démographiques respectives et les événements détaillés. (Les colonnes du fichier texte correspondent aux variables suivantes (NA : Non disponible) : ID du participant, âge, sexe, adulte BMI (kg/m<sup>2</sup>), poids de l'enfant (kg), taille de l'enfant (cm)).

Les publications utilisant cette base de données doivent citer l'article suivant afin d'identifier la base de données :

Rocha BM et al. (2019) " An open access database for the evaluation of respiratory sound classification algorithms " *Physiological Measurement* 40 035001 [208].

## **Annexe 2 - Base de données des Effets Sonores de la BBC**

---

<https://sound-effects.bbcrewind.co.uk/about>

### **À PROPOS DES EFFETS SONORES DE LA BBC**

La BBC Sound Archive est une collection d'enregistrements audio gérée par la BBC et fondée en 1936. Ses enregistrements remontent à la fin du XIXe siècle et comprennent de nombreux éléments rares, notamment des discours contemporains de personnalités publiques et politiques, de la musique folklorique, des dialectes britanniques et des effets sonores [215].

Les archives d'effets sonores de la BBC sont disponibles à des fins personnelles, éducatives ou de recherche. On y trouve plus de 33 000 clips provenant du monde entier et datant des 100 dernières années. Il s'agit notamment de clips réalisés par l'atelier radiophonique de la BBC, d'enregistrements du Blitz de Londres, d'effets spéciaux réalisés pour des productions télévisées et radiophoniques de la BBC, ainsi que de plus de 15 000 enregistrements provenant des archives de la Natural History Unit. Vous pouvez explorer les sons de tous les continents - des cloches des collèges d'Oxford à une chute d'eau en Patagonie - ou écouter le klaxon d'un sous-marin ou le son d'une porte claquée d'une Ford Cortina 1969 [224].

### **LA COLLECTION DE L'UNITÉ D'HISTOIRE NATURELLE**

La collection de sons de l'Unité d'histoire naturelle est née de la nécessité de fournir des sons naturels authentiques pour les productions télévisées et radiophoniques de l'Unité d'histoire naturelle [224], dont la production était en plein essor dans les années 1960. La plupart des gens ne réalisent probablement pas que la majorité des émissions télévisées d'histoire naturelle sont enregistrées en mode muet ; les exigences en matière d'enregistrement des caméras et des microphones étant très différentes ; et généralement, il faut se rapprocher beaucoup plus de son sujet pour enregistrer le son d'un animal que pour le filmer ; ainsi, la plupart du temps, le son est ajouté au cours du montage d'une production. Cela signifie que l'Unité devait constituer une bibliothèque de sons naturels de bonne qualité qui deviendrait une ressource permanente et comblerait les lacunes des productions lorsqu'elles n'étaient pas en mesure de les enregistrer de manière convaincante au moment du tournage.

L'Unité a commencé à assimiler et à incorporer les sons recueillis par les preneurs de son des productions à leur retour du terrain, tout en employant son propre preneur de son pour recueillir les sons naturels et en commandant des enregistrements à d'autres. Petit à petit, la collection de sons s'est constituée et est devenue l'une des plus grandes collections de sons naturels du pays.

La collection d'origine était conservée sur vinyle, avant d'être remplacée par des bandes d'un quart de pouce, puis par des CD, et enfin par une base numérique. Le son ne vieillit pas de la même manière que les films, de sorte que les enregistrements réalisés sur bande dans les années 1970 peuvent encore être utilisés dans des productions de haute qualité aujourd'hui.

La collection a été constituée en grande partie entre les années 1960 et le début des années 2000, date à laquelle les exigences des productions et les avantages de l'enregistrement numérique l'ont emporté sur la nécessité de développer davantage la collection.

L'enregistrement le plus ancien de la collection est probablement le premier enregistrement sonore d'histoire naturelle au monde, réalisé sur cylindre de cire en 1889 par Ludwig Koch et légué par la suite à la BBC. Certains des premiers enregistrements de terrain de la BBC ont été réalisés par David Attenborough, d'autres par des enregistreurs tels que David Toombs et Nigel Tucker dans les années 1980, et Chris Watson ces dernières années.

La collection présente véritablement un éventail mondial de sujets, dont la plupart peuvent être divisés en enregistrements d'espèces spécifiques et en enregistrements d'atmosphères naturelles, bien qu'il y ait également divers enregistrements de phénomènes naturels, ainsi que des chants et de la musique tribaux.

## Annexe 3 – Base de données des sons de battements cardiaques

---

### QU'Y A-T-IL DANS UN BATTEMENT DE CŒUR ?

Cet ensemble de données [220] a été créé à l'origine pour un défi d'apprentissage automatique visant à classer les sons de battements cardiaques. Les données ont été recueillies auprès de deux sources : (A) du grand public via l'application iPhone iStethoscope Pro, et (B) d'un essai clinique dans des hôpitaux utilisant le stéthoscope numérique DigiScope.

### CONTENU

L'ensemble de données est divisé en deux sources, A et B :

1. **Source A :**

- **set\_a.csv** - Étiquettes et métadonnées des battements de cœur collectés auprès du grand public via une application iPhone.
- **setatiming.csv** - contient des informations de synchronisation de référence pour les enregistrements "normaux" de l'ensemble A.

2. **Source B :**

- **set\_b.csv** - Étiquettes et métadonnées des battements cardiaques recueillis lors d'un essai clinique dans des hôpitaux à l'aide d'un stéthoscope numérique.

3. **fichiers audio** - Longueurs variables, entre 1 seconde et 30 secondes. (Certains ont été coupés pour réduire le bruit excessif et fournir le fragment le plus important du son).

### DEFI D'APPRENTISSAGE AUTOMATIQUE DE PETER BENTLEY ET AL.

L'apprentissage automatique consiste à séparer automatiquement les battements de cœur normaux des battements de cœur anormaux et des souffles cardiaques grâce à ce défi d'apprentissage automatique de Peter Bentley et al [225]. L'objectif de la tâche était d'abord (1) d'identifier les emplacements des sons cardiaques à partir de

l'audio, et (2) de classer les sons cardiaques dans l'une de plusieurs catégories (sons de battements de cœur normaux contre divers sons non normaux). Dans ce qui suit nous présentons la description de ce défi d'apprentissage automatique de Peter Bentley et al [225] :

## **Contexte**

Selon l'Organisation mondiale de la santé, les maladies cardiovasculaires (MCV) sont la première cause de décès dans le monde : le nombre de personnes qui meurent chaque année de MCV est supérieur à celui de toute autre cause. On estime que 17,1 millions de personnes sont mortes de MCV en 2004, ce qui représente 29 % de tous les décès dans le monde. Parmi ces décès, on estime que 7,2 millions étaient dus à des maladies coronariennes. Toute méthode permettant de détecter les signes de maladie cardiaque pourrait donc avoir un impact significatif sur la santé mondiale. Ce défi consiste à produire des méthodes permettant d'atteindre cet objectif. Plus précisément, nous souhaitons créer le premier niveau de dépistage des pathologies cardiaques à la fois dans un environnement hospitalier par un médecin (à l'aide d'un stéthoscope numérique) et à domicile par le patient (à l'aide d'un appareil mobile).

Le problème est d'un intérêt particulier pour les chercheurs en apprentissage automatique car il implique la classification de données d'échantillons audio, où la distinction entre les classes d'intérêt n'est pas triviale. Les données sont recueillies dans des situations réelles et contiennent fréquemment des bruits de fond de tous types imaginables. Les différences entre les sons cardiaques correspondant à différents symptômes cardiaques peuvent également être extrêmement subtiles et difficiles à distinguer. Pour réussir à classer ce type de données, il faut des classificateurs extrêmement robustes. Malgré son importance médicale, il s'agit à ce jour d'une application relativement inexplorée de l'apprentissage automatique.

## **Aperçu de la tâche**

Les données ont été recueillies à partir de deux sources : (A) auprès du grand public via l'application iPhone iStethoscope Pro, fournie dans l'ensemble de données A, et (B) à partir d'un essai clinique dans des hôpitaux utilisant le stéthoscope numérique DigiScope, fourni dans l'ensemble de données B.

### *DÉFI 1 - Segmentation du son du cœur*

Le premier défi consiste à produire une méthode capable de localiser les sons S1(lub) et S2(dub) dans les données audio, en segmentant les fichiers audio normaux dans les deux ensembles de données. Pour permettre à votre méthode d'apprentissage automatique d'apprendre, nous fournissons l'emplacement exact des sons S1 et S2 pour certains des fichiers audio. Vous devez les utiliser pour identifier et localiser les sons S1 et S2 de tous les battements de cœur dans le groupe non étiqueté. Les emplacements des sons sont mesurés en échantillons audio pour une meilleure précision. Votre méthode doit utiliser la même unité.

### *DÉFI 2 - Classification des bruits du cœur*

La tâche consiste à produire une méthode capable de classer les sons cardiaques réels (également appelée "classification des battements") dans l'une des quatre catégories de l'ensemble de données A :

1. Normal
2. Murmure
3. Son cardiaque supplémentaire
4. Artifact

Et trois classes pour le jeu de données B :

1. Normal
2. Murmure
3. Extrasystole

### **Téléchargements**

Il y a deux ensembles de données :

1. *Le jeu de données A, contenant 176 fichiers au format WAV, organisés comme suit*

Atraining\_normal.zip 14Mb 31 fichiers à télécharger

Atraining\_murmur.zip 17.3Mb 34 fichiers à télécharger

Atraining\_extrahs.zip 6.9Mb 19 fichiers à télécharger

Atraining\_artifact.zip 22.5Mb 40 fichiers à télécharger

Aunlabelledtest.zip 24.6Mb 52 fichiers à télécharger

Les mêmes ensembles de données sont également disponibles au format aif :

Atraining\_normal.zip 13.2Mb 31 fichiers à télécharger

Atraining\_murmur.zip 16.4Mb 34 fichiers à télécharger

Atraining\_extrahs.zip 6.5Mb 19 fichiers à télécharger

Atraining\_artifact.zip 20.9Mb 40 fichiers à télécharger

Aunlabelledtest.zip 23.0Mb 52 fichiers à télécharger

Données de segmentation (mises à jour le 23 mars 2012), donnant les emplacements des sons S1 et S2 dans Atraining\_normal :

Atraining\_normal\_seg.csv

2. *Ensemble de données B, contenant 656 fichiers au format WAV, organisé comme suit :*

Btraining\_normal.zip (contenant le sous-répertoire Btraining\_noisynormal)  
13.8Mb 320 fichiers à télécharger



Btraining\_murmur.zip (contenant le sous-répertoire Btraining\_noisymurmur)  
5.3Mb 95 fichiers à télécharger

Btraining\_extrasystole.zip 1.9Mb 46 fichiers à télécharger

Bunlabelledtest.zip 9.2Mb 195 fichiers à télécharger

Les mêmes jeux de données sont également disponibles au format aif :

Btraining\_normal.zip (contenant le sous-répertoire Btraining\_noisynormal)  
13.0Mb 320 fichiers à télécharger

Btraining\_murmur.zip (contenant le sous-répertoire Btraining\_noisymurmur)  
5.1Mb 95 fichiers à télécharger

Btraining\_extrasystole.zip 2.1Mb 46 fichiers à télécharger

Bunlabelledtest.zip 8.7Mb 195 fichiers à télécharger

Données de segmentation, donnant les emplacements des sons S1 et S2 dans  
Btraining\_normal : Btraining\_normal\_seg.csv

Les deux défis qui étaient associés à ce concours :

### **1. Segmentation du son du cœur**

Le premier défi est de produire une méthode qui peut localiser les sons S1(lub) et S2(dub) dans les données audio, en segmentant les fichiers audio normaux dans les deux ensembles de données (segmenter les fichiers audio dans Atraining\_normal.zip et Btraining\_normal.zip en utilisant les segmentations d'entraînement fournies ci-dessus).

### **2. Classification des sons cardiaques**

La tâche consiste à produire une méthode capable de classer les sons cardiaques réels (également connus sous le nom de "classification des battements") dans l'une des quatre catégories qui vont suivre (étiqueter correctement les sons dans Aunlabelledtest.zip et Bunlabelledtest.zip).

## **Description et organisation des données**

Il faut utiliser la citation suivante si les données sont utilisées [225] :

```
@misc{pascal-chsc-2011, author = "Bentley, P. and Nordehn, G. and  
Coimbra, M. and Mannor, S.", title = "Le projet {PASCAL} {Défi {C}de  
classification des {S}sonorités {C} 2011 {(CHSC2011)} {R}résultats",  
howpublished = "http://www.peterjbentley.com/heartchallenge/index.html" }
```

Les fichiers audio sont de longueurs variables, entre 1 seconde et 30 secondes (certains ont été coupés pour réduire le bruit excessif et fournir le fragment le plus important du son).

La plupart des informations contenues dans les sons cardiaques se trouvent dans les composantes de basse fréquence, avec du bruit dans les fréquences plus élevées. Il est courant d'appliquer un filtre passe-bas à 195 Hz. Les transformées de Fourier rapides sont également susceptibles de fournir des informations utiles sur le volume et la fréquence dans le temps. Des connaissances plus spécifiques au domaine concernant la différence entre les catégories de sons sont fournies ci-dessous.

### ***Catégorie normale***

Dans la catégorie Normal, on trouve des bruits cardiaques normaux et sains. Ils peuvent contenir du bruit dans la dernière seconde de l'enregistrement, lorsque le dispositif est retiré du corps. Ils peuvent contenir une variété de bruits de fond (de la circulation aux radios). Ils peuvent également contenir des bruits aléatoires occasionnels correspondant à la respiration, ou au frottement du microphone contre les vêtements ou la peau. Un bruit cardiaque normal présente un schéma clair "lub dub, lub dub", le temps entre "lub" et "dub" étant plus court que le temps entre "dub" et le "lub" suivant (lorsque la fréquence cardiaque est inférieure à 140 battements par minute). Notez la description temporelle des emplacements "lub" et "dub" dans le temps dans l'illustration suivante :

...lub.....dub..... lub.....dub..... lub.....dub..... lub.....dub...

En médecine, nous appelons le son lub "S1" et le son dub "S2". La plupart des fréquences cardiaques normales au repos se situent entre 60 et 100 battements ('lub dub's) par minute. Cependant, notez que, puisque les données peuvent avoir été recueillies auprès d'enfants ou d'adultes dans des états calmes ou excités, les fréquences cardiaques dans les données peuvent varier de 40 à 140 battements ou plus par minute. L'ensemble de données B contient également des données noisy\_normal - des données normales qui comprennent une quantité importante de bruit de fond ou de distorsion. Vous pouvez choisir de les utiliser ou de les ignorer, mais l'ensemble de test comprendra des exemples tout aussi bruyants.

### ***Catégorie de murmure***

Les souffles cardiaques ressemblent à un bruit de "sifflement, de grondement ou de fluide turbulent" à l'un des deux emplacements temporels suivants : (1) entre "lub" et "dub", ou (2) entre "dub" et "lub". Ils peuvent être le symptôme de nombreux troubles cardiaques, parfois graves. Il y aura toujours un "lub" et un "dub". L'une des choses qui confondent les personnes n'ayant pas de formation médicale est que les souffles se produisent entre lub et dub ou entre dub et lub ; pas sur lub et pas sur dub. Ci-dessous, vous trouverez un astérisque\* aux endroits où un murmure peut se trouver.

...lub..\*\*\*..dub..... lub..\*\*\*..dub ..... lub..\*\*\*..dub ..... lub..\*\*\*..dub ...

ou

...lub.....dub...\*\*\*\*\*...lub..... dub...\*\*\*\*\*...lub ..... dub...\*\*\*\*\*...lub .....dub...

L'ensemble de données B contient également des données "noisy\_murmur" - des données murmure qui comprennent une quantité importante de bruit de fond ou de distorsion. Vous pouvez choisir de les utiliser ou de les ignorer, mais l'ensemble de test comprendra des exemples tout aussi bruyants.

***Catégorie des sons cardiaques supplémentaires (Ensemble de données A)***

Les bruits cardiaques supplémentaires peuvent être identifiés parce qu'il y a un son supplémentaire, par exemple un "lub-lub dub" ou un "lub dub-dub". Un bruit cardiaque supplémentaire peut ne pas être un signe de maladie. Cependant, dans certaines situations, il s'agit d'un signe important de maladie qui, s'il est détecté à temps, peut aider la personne. Il est important de pouvoir détecter le bruit cardiaque supplémentaire car il n'est pas très bien détecté par l'échographie. Ci-dessous, notez la description temporelle des bruits cardiaques supplémentaires :

...lub.lub.....dub..... lub.lub.....dub.....lub.lub.....dub.....

ou

...lub..... dub.dub.....lub.....dub.dub.....lub.....dub. dub.....

***Catégorie Artefact (Ensemble de données A)***

Dans la catégorie des artefacts, on trouve un large éventail de sons différents, notamment des crissements et des échos de larsens, des paroles, de la musique et des bruits. Il n'y a généralement pas de sons cardiaques discernables, et donc peu ou pas de périodicité temporelle à des fréquences inférieures à 195 Hz. Cette catégorie est la plus différente des autres. Il est important de pouvoir la distinguer des trois autres catégories, afin de pouvoir demander à la personne qui recueille les données de réessayer.

***Catégorie Extrasystole (ensemble de données B)***

Les sons d'extrasystole peuvent apparaître occasionnellement et peuvent être identifiés parce qu'il y a un son cardiaque qui n'est pas rythmé, avec des battements cardiaques supplémentaires ou sautés, par exemple un "lub-lub dub" ou un "lub dub-dub". (Ce n'est pas la même chose qu'un bruit cardiaque supplémentaire car l'événement ne se produit pas régulièrement). Une extrasystole peut ne pas être un signe de maladie. Elle peut se produire normalement chez un adulte et peut être très fréquente chez les enfants. Cependant, dans certaines situations, les extrasystoles peuvent être causées par des maladies cardiaques. Si ces maladies sont détectées plus tôt, le traitement sera probablement plus efficace. Ci-dessous, notez la description temporelle des extrasystoles cardiaques :

.....lub.....dub.....lub.....dub.....lub.lub.....dub.....

ou

...lub.....dub.....lub.....dub.dub.....lub.....dub.....