デジタル音楽への変形離散コサイン変換と差分拡散法による電子透かし

岩切宗利† 松井甲子雄†

この論文では，時間周波数変換と統計的分布の拡散を用いて高品質なデジタル音楽の著作権を保護するための電子透かし法を提案する．その原理は，音楽データを変形離散コサイン変換（MDCT）を用いて周波数係数値へ変換し，その特定成分をランダムに制御することにより透かしを埋め込むものである．すなわち，ランダムな系列として準備した鍵乱数によって選択された2つのMDCT係数を制御し，それらから得られる統計的な分布の偏りを拡散するものである．この鍵データを所有しない者には，埋込みのあるMDCT係数の組を特定できないため第3者に対して透かしの存在を秘匿できる．さらに，帯域通過フィルタにより音声を処理されてもその影響を受けずに透かしが残存する．本手法を用いた実験では，40 dB以上の高音質を維持しつつ，MPEGオーディオなどの高能率圧縮やJitter Attackを施しても透かしが消失しないことを確かめている．

Digital Watermarking Scheme to Audio Data by Spread Differential Method and Modified Discrete Cosine Transform

Munetoshi Iwakiri† and Kineo Matsui†

This paper presents a digital watermark scheme for high quality audio data. The scheme embeds watermarks in the form of randomly controlled perterbation over the statistical distribution of an MDCT (Modified Discrete Cosine Transform) sequence. A pair of watermark signals is uniformly spread over the whole window frame when it is embedded to sound, and the watermarked noise is suppressed to very low level. Our scheme achieves a sound quality of over 40 dB in segmental SN ratio. The differential spreading scheme on MDCT coefficients brings us a robust watermark for copyright protection, thus the watermark is proof against such attacks as bandpass filtering, amplification, lossy compression and Jitter Attack.

1. はじめに

音楽データは，アナログの音声波形をサンプリング定理に基づいて標本化し，量子化して線形パルス符号化（PCM：Pulse Code Modulation）[1]することにより生成されている．特に，CDなどの高音質データは，サンプリングレート44.1 kHz，16 bit量子化のPCMによりデジタル化されている．この方法によると，人間の可聴周波数帯域をすべてカバーできるため，聴感的に高い音質を保ったままデータ化できる．

一方，このデジタルデータは完全な状態のまま容易に複製できる特徴がある．この忠実な再現性はデジタル化の大きな利点である反面，デジタル著作物の著作権保護が必須の要件になる．この対策として，人間が知覚できないように著作権情報を音楽データへ埋め込む電子透かし[2]が注目されている．

音楽コンテンツへの電子透かしやそれに類する試みとして，Boneyらによる聴感的マスキングを利用する手法[3]，松井らの量子化雑音に見せかける手法[4]，岩切らの圧縮音声符号に直接埋め込む手法[5)~7]や直接拡散法[8]，富岡らの音源定位制御法[9]および松本らのシンセサイザ符号（SMF）への埋込み法[10]などが検討されている．これらの手法によれば，聴感的な音質をほとんど劣化することなく透かし信号を埋め込むことができる．特に，直接拡散[11]を利用する電子透かし技術[8]は，狭帯域な透かし信号の時系列標本値をランダム変調し，広い帯域全体にわたって分散配置（拡散）するため，拡散符号列（擬似乱数）を秘密鍵として高い秘匿性を実現できる．また，デジタル画像を対象とした電子透かしとして，パッチワークとよばれる手法[12]が提案されている．これは，ランダムに選択した2つの時系列標本値を制御し，それらの値から得られる分布のピーク位置をずらす手法である．この方法

† 防衛大学校情報工学科
Department of Computer Science, National Defense Academy

によれば，標本値の選択性をランダムにできるため直接拡散法と同様に高い秘匿性を実現できる．

しかし，直接拡散法やパッチワーク法のように時系列標本値をランダムに制御する手法は，透かしを検出する際の乱数同期が正確でなければならない．これは高い秘匿性を実現できる反面，電子透かしへの攻撃手法として知られる Jitter Attack [13]のような信号同期を崩す攻撃に脆弱である．本研究では，時系列上での微少な変調による影響を受けにくい周波数領域に透かしを埋め込むことにより攻撃耐性の向上を試みた．

その方法は，まず各音声フレームごとに周波数変換して得られる係数の中から，各コンテンツごとに異なる乱数列を用いて選ばれた周波数係数を制御し，統計的な手法に基づく埋込みを施すものである．この単位時間あたりの埋込み量を多くすれば，オリジナルの音楽データの統計的分布を変化させることができる．透かしを検出する際は，検査対象とする音楽データの周波数成分を調べ，統計的に透かしの有無を判定すればよい．本方式では，埋込みを施した周波数成分の位置を検出鍵として高い秘匿性を実現できる．また，時系列から周波数係数への変換に変形離散コサイン変換（MDCT）[14]を用いることにより，音声フレーム間に生じやすいスパイク雑音や歪みを低減した．提案方式は，従来の直接拡散法[8]のように狭帯域の透かし信号を時系列上での拡散により広い帯域へ分散配置する技術や，時系列そのものを埋込み操作するパッチワーク法[12]とは本質的に異なる原理に基づく技術である．

2 章では変形離散コサイン変換と統計的電子透かしによる埋込みの手法を示す．3 章では本方式を原理とする埋込みの特性について考察する．4 章では，提案方式に基づいた実験システムを構成し，2，3 の性能評価を行った結果について示す．

2. 透かしの埋込みと検出

2.1 提案方式の原理

本方式では，ある系列から得られる統計的な分布をランダムに制御し，その分布の変化を透かし信号として検出する手法を用いた．本原理の概要を**図 1** に示す．まず，M 個の標本値により構成される $i(=0,1,2,\cdots)$ 番目の音声フレームから離散周波数係数 $X_i(k), 0 \leq k \leq M-1$ を求める．次に，その中からランダムに 2 つの係数 $X_i(a_i)$ および $X_i(b_i)$ を選び，その差分値

$$d_i = X_i(a_i) - X_i(b_i) \qquad (1)$$
$$0 \leq a_i \leq M-1,\ 0 \leq b_i \leq M-1$$

を算出する．一方，各周波数係数値 $X_i(a_i)$，$X_i(b_i)$ それぞれに対し，乱数 u_i，v_i を加算したものを $X'_i(a_i)$，$X'_i(b_i)$ とする．これらの差分値は

$$\begin{aligned} d'_i &= X'_i(a_i) - X'_i(b_i) \\ &= X_i(a_i) + u_i - \{X_i(b_i) + v_i\} \\ &= d_i + u_i - v_i \end{aligned} \qquad (2)$$

になる．この導出から d'_i は，d_i がランダムな値の加減による影響を受けたものであると見なせる．よって，埋込みのない状態の差分値 d_i の統計的分布に比べると，d'_i の分布は，よりランダム性が増したものになる．すなわち，ランダムな埋込みが施された 2 つの係数を抽出し，その差分統計を調べると埋込みのない場合と異なる特徴的な分布が得られることになる．本論文では，この原理に基づいた周波数係数差分値の統計的分布を拡散する電子透かし技術について述べる．

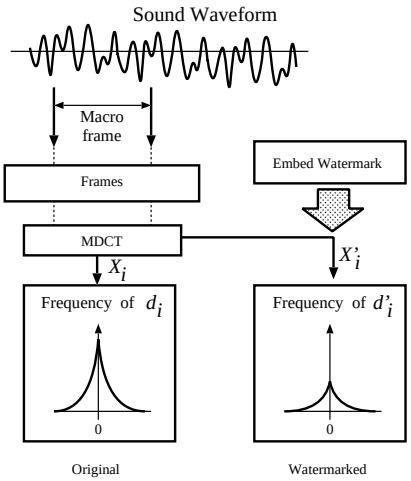


図 1 埋込みの原理
Fig. 1 Watermark embedding rule.

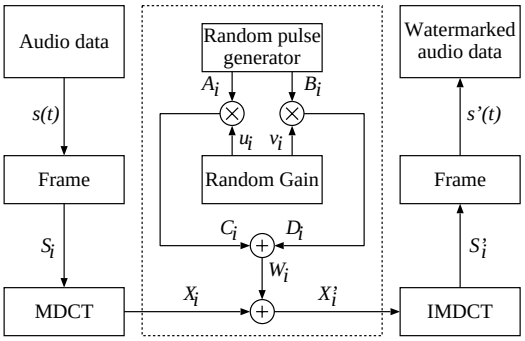


図 2 埋込みシステム
Fig. 2 Watermarking system.

2.2 埋込み法

本手法による電子透かしの処理ブロックを**図 2** に示す．その方法では，まず音楽データから抽出した音声信号を一定長の音声フレームに区切り，変形離散コサイン変換（MDCT）する．MDCT [14]は，**図 3** に示

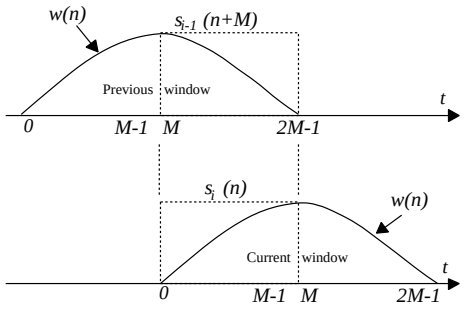


図 3 MDCT の窓関数
Fig. 3 MDCT window to signal interference.

すように M 個のスペクトル係数を求めるために $2M$ 個の時系列サンプルを用いる．これは周波数分離度を高くし，かつ，隣接するフレームを互いに重複させることによりフレーム間歪みを抑制できる手法である．

サンプリング時刻 t における音声信号を $s(t)$ としたとき，$i(=0,1,2,\cdots)$ 番目の音声フレーム S_i の MDCT 係数 $X_i(k)$ は

$$X_i(k) = \frac{2}{M}\sum_{n=0}^{2M-1} w(n)\cdot c(k,n)\cdot s(n+iM) \quad 0\le k\le M-1,\ 0\le n\le 2M-1 \tag{3}$$

により求まる．窓関数 $w(n)$ および MDCT 基底 $c(k,n)$ は，それぞれ

$$w(n) = \sin\left(\frac{\pi(2n+1)}{4M}\right) \tag{4}$$

$$c(k,n) = \cos\left(\frac{\pi(2k+1)(2n+M+1)}{4M}\right) \quad 0\le k\le M-1,\ 0\le n\le 2M-1 \tag{5}$$

である．ここで，乱数列 R（検出鍵）から音声フレーム S_i ごとに異なる値 $\{a_i : 0\le a_i\le M-1\}$ を抽出し，

$$A_i(k) = \begin{cases} 1, & if\ k = a_i \\ 0, & if\ k \ne a_i \end{cases} \quad 0\le k\le M-1 \tag{6}$$

として M 個の要素 $A_i(k)$ を持つ系列 A_i を生成する．たとえば，$M=10$ のとき乱数 $a_i=3$ ならば，

$$A_i = \{0,0,0,1,0,0,0,0,0,0\} \tag{7}$$

となる．さらに，A_i の各系列値に乱数 $\{u_i : |u_i|\le p\}$ を乗じて C_i を生成する．ただし，p は埋込みの強度を制御するパラメータである．先に示した A_i に乱数 $u_i=5$ を適用すると

$$C_i = \{0,0,0,5,0,0,0,0,0,0\} \tag{8}$$

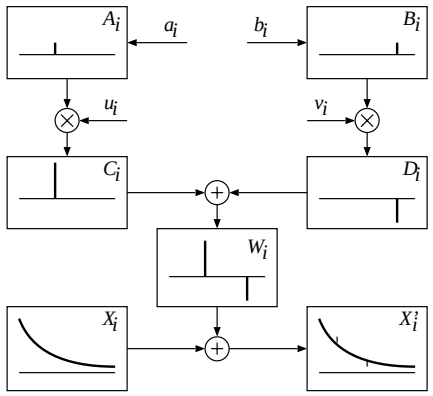


図 4 埋込み手順
Fig. 4 Watermark embedding scheme.

が生成されることになる．同様の手順を用いて鍵乱数列 R から抽出した $\{b_i : 0\le b_i\le M-1\}$ により B_i を生成し，これに乱数 $\{v_i : |v_i|\le p\}$ を乗じて D_i を合成する．たとえば，$M=10$，$b_i=7$，$v_i=-3$ のとき，

$$D_i = \{0,0,0,0,0,0,0,-3,0,0\} \tag{9}$$

が得られる．次に，C_i と D_i の各系列を加算し，埋込み制御系列 W_i を得る．ここに述べた例の場合

$$W_i = \{0,0,0,5,0,0,0,-3,0,0\} \tag{10}$$

となる．この埋込み系列 W_i の各要素 $W_i(k)$ と $X_i(k)$ を加算し埋込み済み MDCT 係数

$$X_i'(k) = X_i(k) + W_i(k) \quad 0\le k\le M-1 \tag{11}$$

を得る（**図 4** 参照）．この処理を各フレームごとに再帰的に f 回だけ埋込みを施し，得られた $X_i'(k)$ を逆変換することにより埋込み済みの第 $i(=0,1,2,\cdots)$ 音声フレームを生成する．ここで，f は，各音声フレームに対して透かし信号を埋め込む頻度を制御するパラメータである．この手順による埋込みを複数の音声フレームへ施し，埋込み済み音声データを生成する．

2.3 透かしの検出

埋込みに用いたものと同じ鍵乱数列 R から，a_i を抽出し式 (6) を用いて M 個の値を持つ系列 A_i を生成する．一方，i 番目の音声フレームから得られた $2M$ サンプルの標本値 S_i' を MDCT 係数 $X_i'(k)$ へ変換する．ここで，A_i の各要素 $A_i(k)$ と $X_i'(k)$ を畳み込み積分すると

$$X_i'(a_i) = \sum_{k=0}^{M-1} X_i'(k)\cdot A_i(k) \tag{12}$$

が得られる．また，同様の手順により R から b_i を抽出し，

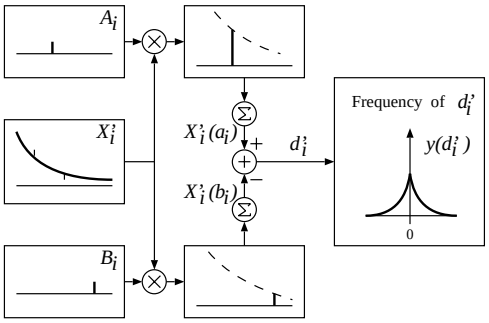


図 5 検出手順
Fig. 5 Watermark detection scheme.

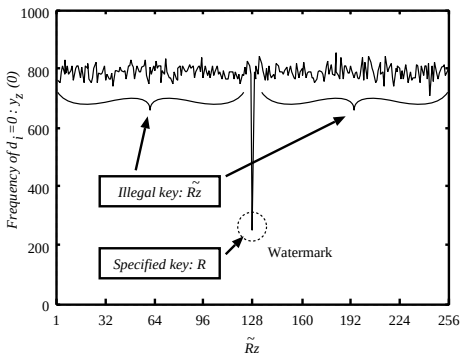


図 6 透かし信号検出の様子
Fig. 6 An example of detection process.

$$X'_i(b_i) = \sum_{k=0}^{M-1} X'_i(k) \cdot B_i(k) \tag{13}$$

を求める．ここで，$X'_i(a_i)$ と $X'_i(b_i)$ の差分値

$$d'_i = X'_i(a_i) - X'_i(b_i) \tag{14}$$

を計算する．このとき，d_i を最近傍の整数値へ丸め，各値の出現頻度を計数値

$$y(d_i) \leftarrow y(d_i) + 1 \tag{15}$$

により調べる．この抽出処理ブロックを**図 5** に示す．各音声フレーム S'_i ごとに埋込みが施された f 組の周波数係数を用いて検査し統計的な分布を得る．

一方，埋込みに用いた鍵乱数列 R とは異なる乱数列 $\tilde{R}_z(z = 1, 2, \cdots)$ を複数準備する．その各組ごとに前述の手順を用いて，その差分値 d_i を求め出現頻度 $y_z(d_i)$ を調べる．この統計的な分布の検査を擬似乱数鍵 $\tilde{R}_z$ について個々に行う．その結果 $\tilde{R}_z(z = 1, 2, \cdots)$ が鍵系列 R と異なるため，埋込みの施されていない周波数成分を用いて d_i を計算するので，その頻度はもとの音声データから得られる分布とほぼ同じものになる．

そこで，$y(d'_i)$ および $y_z(d_i)(z = 1, 2, \cdots)$ の分布から差分値が 0 の頻度すなわち $y_z(0)$ のみを抽出しグラフ化すると**図 6** になる．図 6 から鍵乱数列 R による検出頻度が，他の非鍵乱数列 $\tilde{R}_z(z = 1, 2, \cdots)$ による頻度より低い値になる事実により透かしを証明できる．

3. 提案方式の特徴

3.1 透かしの秘匿性

ここで提案する手法は，各音声フレームごとに異なる複数組の周波数係数へ埋込みを施すものである．したがって，本方式による透かしを不正に抽出するには，その周波数の組を確実に特定できなければならない．たとえば，ある音声フレームから M 個の周波数係数が生成されたとき，埋込み成分の組合せは M^2 通りある．よって，I 個の音声フレームそれぞれに一組の透かしが埋め込まれているとき，それを抽出するために必要な乱数 (a_i, b_i) の組合せは M^{2I} である．たとえば，$M = 1024$，$I = 50$ のときの組合せ総数は約 10^{300} になる．よって，埋込みに用いた鍵乱数列 R を知らない第 3 者にとって，透かしが埋め込まれた周波数成分を正確に特定するのは困難である．したがって，本方式は高い秘匿性を実現できる電子透かし技術の 1 つと見なせる．

3.2 帯域通過フィルタ耐性

本方式による透かし信号は，2 つの MDCT 係数の組として帯域全体に存在する．したがって，帯域通過フィルタによって攻撃を受けてもその影響を受ける音声区間は限定される．音質を高く維持するためには，一般に通過帯域幅を制限帯域に比して広く設定する．よって，本方式によれば大部分の透かし成分が帯域フィルタによる影響を受けずにすむ．ただし，組となる MDCT 成分が通過制限帯域にともに含まれるとその差分値は 0 に近い値となるため，埋込み鍵 R を用いた検出 $y(0)$ とそれ以外の鍵 $\tilde{R}_z$ を用いて得られた $y_z(0)$ の差が少なくなると予想される．したがって，音質の劣化と引き換えに通過帯域幅を狭く設定して攻撃されると検出が難しくなる．ただし，周波数成分の強さから通過制限された帯域を判別できれば，そのような攻撃にも対処できる．

3.3 透かしの多重化

本方式による埋込みはランダムに選んだ周波数係数組の差分値の分布 y の偏りを拡散するものである．よって，同一のコンテンツに対して，多重に埋込みを施しても埋込みに用いた周波数係数組 (a_i, b_i) が異なれば，お互いに影響を及ぼすことはない．偶然に同じ周波数係数の組へ埋込みを施しても，それらの差分値に加えられた値はランダムであるため完全に消失する

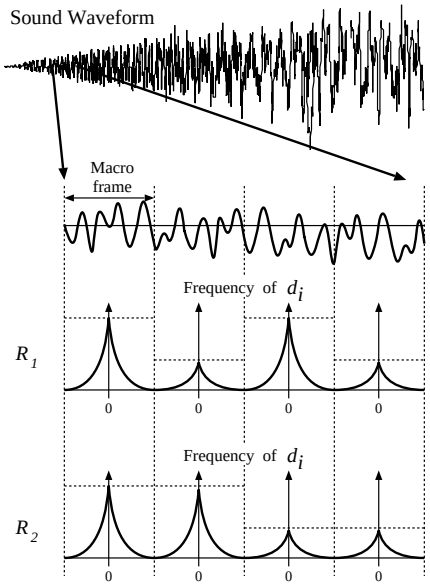


図 7 透かしの多重埋込み
Fig. 7 Multiplex watermarking.

ことは少ない．したがって，埋込みの一部が重複しても，透かしの検出結果に与える影響はほとんどないと考えられる．この特徴を利用すれば，**図 7** のように複数のビット情報を矩形波として埋め込むこともできる．この図では，同一のコンテンツに 2 種類の鍵乱数列 R_1，R_2 を用いて透かしを埋め込む際に，複数の MDCT 変換フレームから成るマクロフレームごとに埋込みの有無を制御した一例である．ただし，多重に透かしを埋め込むと，それだけ音質に与える影響も増大するので注意しなければならない．

4. 実験結果と考察

CD などの高音質な音楽データに電子透かしを埋め込む際，音質の劣化が少なく，さらにデータ量を高能率圧縮しても透かしが消失しにくいことが望ましい．本実験では，高品質な音楽データに埋込みを施したとき，音質に与える影響および MPEG1 Audio LayerIII（MP3）[15]による高能率圧縮などのデジタル信号処理が透かしに与える影響について検討した．

4.1 準 備

表 1 の実験データは，音楽 CD の再生音をサンプリングレート 44.1 kHz，16 bit 量子化の条件によりデジタル化したものである．通常，これらの音楽データはステレオ音であるが，議論を簡単にするため，その片側成分のみを抽出して実験データとした．これらの実験データの周波数スペクトルの分布を**図 8** に示した．Classic はバイオリンによる伸びのある演奏である．Jazz の演奏は，ピアノ音による強いアタックのスペクトルが間断的に発生し，それらが徐々に減衰する特徴がある．Dance は，合成音を含むハイテンポかつ連続的な複合音であり，広い帯域にわたってスペクトルが分布している．また，実験に用いる鍵乱数集合 $\mathcal{R}$ として，あらかじめ 256 種類の疑似乱数列（R_1〜R_{256}）を準備した．これらは通常，疑似乱数発生器に秘密鍵を設定することにより生成できる．

表 1 実験音声
Table 1 Sound for experiment.

	Samples	Sec.
Classic	440,832	10
Jazz	440,832	10
Dance	440,832	10

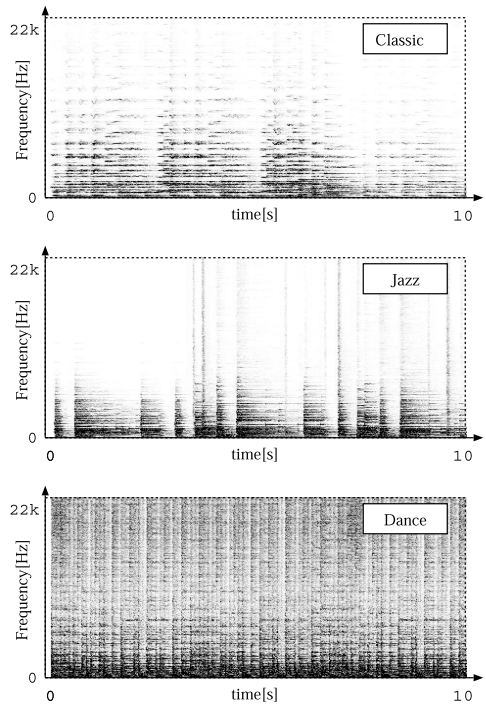


図 8 実験音声の周波数分布
Fig. 8 Spectrograph of sound for experiment.

4.2 音質の評価法

音質の客観的な評価尺度として最も基本的なものに信号対量子化雑音比（SNR：Signal to quantization Noise Ratio）がある．SNR[dB] の評価式は，入力音声 $So(m)$ とその量子化誤差 $Er(m)$ を用いて次のように定義される[1]．

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_m So^2(m)}{\sum_m Er^2(m)} \tag{16}$$

ここでは，SNR を改良して主観評価との対応関係を向上した $\mathrm{SNR}_{\mathrm{seg}}$（Segmental SNR）を用いた[1]．

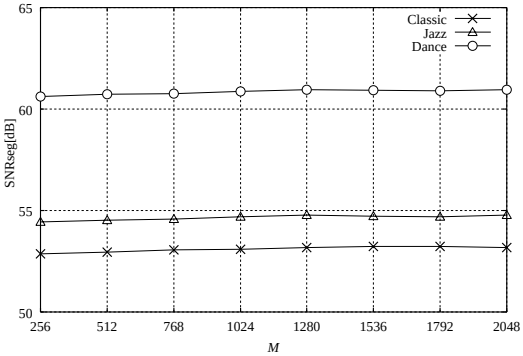


図 9 M に対する $\mathrm{SNR}_{\mathrm{seg}}$（$p=5$，$f=1$）
Fig. 9 $\mathrm{SNR}_{\mathrm{seg}}$ to M ($p=5$, $f=1$).

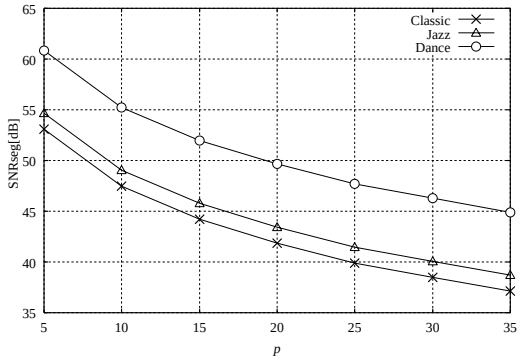


図 10 p に対する $\mathrm{SNR}_{\mathrm{seg}}$（$M=1024$，$f=1$）
Fig. 10 $\mathrm{SNR}_{\mathrm{seg}}$ to p ($M=1024$, $f=1$).

$$\mathrm{SNR}_{\mathrm{seg}} = \frac{1}{N_h}\sum_{h=1}^{N_h}\mathrm{SNR}_{\mathrm{h}} \quad [\mathrm{dB}] \tag{17}$$

N_h は測定区間のフレーム数を表し，$\mathrm{SNR}_{\mathrm{h}}$ は，h フレームにおける SNR である．本実験では，1 フレームの長さを 32 ms とした．また，誤差のない音声フレームすなわち，$\mathrm{SNR}_{\mathrm{h}}=\infty$ の音声フレームを除外して測定した．

4.3 音質への影響

表 1 に示した実験音声へ透かしを埋め込んだときの音質を調べた．まず，埋込みの強度 $p=5$，各フレームへの埋込み頻度 $f=1$ に固定し，(a_i, b_i) の値域を決定する MDCT 変換のフレーム長 M を変化させると，**図 9** が得られた．この結果から M を小さな値として埋込み頻度を増大しても，顕著な音質劣化を生じないことが分かった．これは，埋込み制御の影響が MDCT フレーム全体へ分散されるためである．次に，$M=1024$，$f=1$ として埋込み強度パラメータ p を変化させると**図 10** が得られた．この結果から透かし強度 p が弱いほど音質に与える影響を少なくできることが分かる．これは埋込みに用いるランダムな制御乱数（u_i，v_i）が一般に小さくなることに起因する．さらに，$M=1024$，$p=5$ として各音声フレームへの埋込み頻度 f を変化させ，埋込みが音質に及ぼした影響を調べると**図 11** のようになった．この図から f の値が大きくなるにつれて $\mathrm{SNR}_{\mathrm{seg}}$ が劣化するものの，その低下率は徐々に少なくなることが分かった．

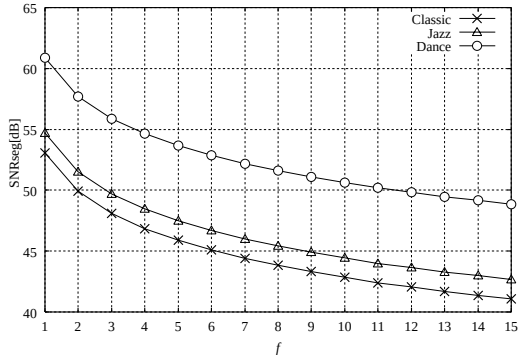
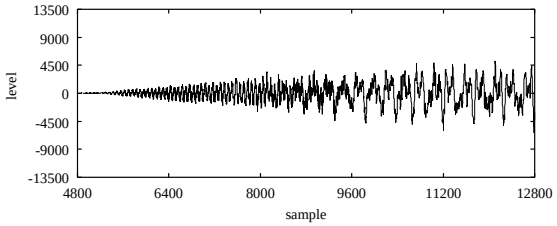


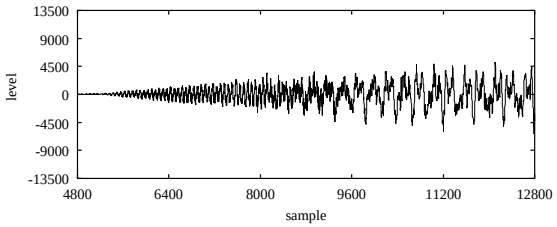
図 11 f に対する $\mathrm{SNR}_{\mathrm{seg}}$（$M=1024$，$p=5$）
Fig. 11 $\mathrm{SNR}_{\mathrm{seg}}$ to f ($M=1024$, $p=5$).

これらの実験結果から p，f の値を大きくすることにより，透かしの存在を顕著にできる反面，音質の劣化量を増大する特性が明らかである．また，特定のパラメータを用いた場合の Dance の音質劣化は，ほかの実験音声に比べて少ないことも分かる．これは，Dance の周波数スペクトルが図 8 に示したように帯域全体にわたって強く分布していることに起因する．よって，提案方式による埋込みが音質に与える影響は，対象音声の周波数スペクトル分布の状態に強く依存するといえる．したがって，埋込みパラメータ p，f は，埋込み対象音声の周波数スペクトル分布が全体的に強い場合には大きな値とし，逆に弱い場合には小さな値を用いることになる．ただし，極端に小さなパラメータ値を用いると，透かし信号の検出が難しくなるため，許容できる音質劣化の範囲を考慮した最大値とすることが望ましい．ここでは，議論を簡単にするために埋込みパラメータを適応化していないが，周波数スペクトル分布と聴覚特性を考慮した適応化も容易である．

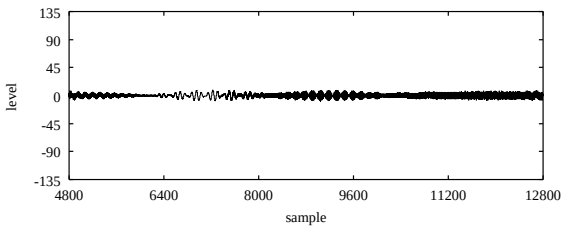
また，本方式では音声フレームごとに異なる周波数成分を操作するため，フレーム間に予測できない歪みを生じることも予想される．MDCT は，符号量圧縮処理などによるフレーム間歪みを発生しにくいとされている．ここでは，**図 12** にオリジナルの波形と $M=1024$，$p=5$，$f=1$ を用いて透かしを埋め込んだ波形の形状を比較し，フレーム間に不自然な不連続点が混入していないことを示す．まず，図 12 (a) に



(a) Decoded waveform without watermark

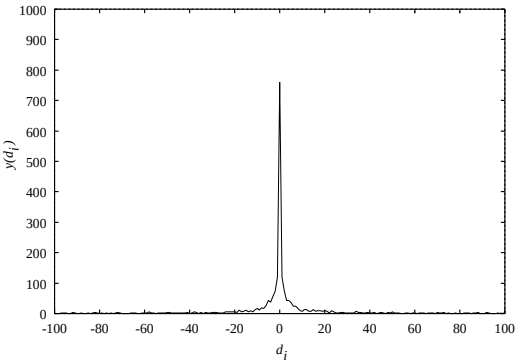


(b) Decoded waveform with watermark

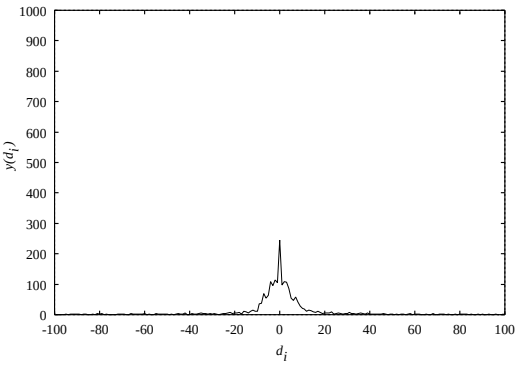


(c) Difference waveform

図 12 音声波形の比較
Fig. 12 Comparison of sound waveform.



(a) Original



(b) Watermarked

図 13 d_i の分布：$y(d_i)$
Fig. 13 Frequency of d_i : $y(d_i)$.

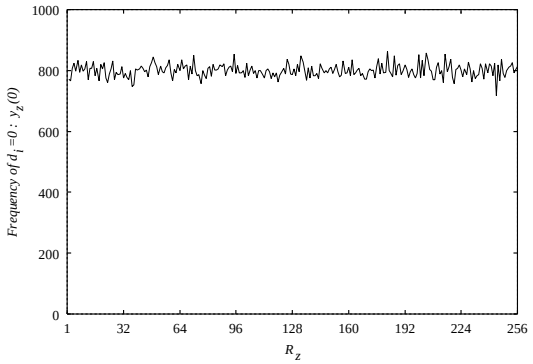
Classic のバイオリン演奏音の一部（8,000 サンプル）を抽出した波形を示し，埋込みを施した同じ音声区間の波形を図 12 (b) のように比較した．さらに，それらの差分波形を図 12 (c) に示すことにより，透かし信号の状態を観察した．ここで，図 12 (c) は，埋込みによる影響が分かりやすいように縦軸を 100 倍に拡大してある．これらの波形から音声フレーム間に不自然な波形歪を生じておらず，透かし信号の波形が自然に変化していることを確認できる．これは埋込みのための周波数変換に，MDCT を用いることにより隣接するフレームを干渉させたためである．また，著者らの聴取では，埋込みを施した再生音声から不自然な歪みを感じられなかった．

4.4 透かしの検出と多重埋込み

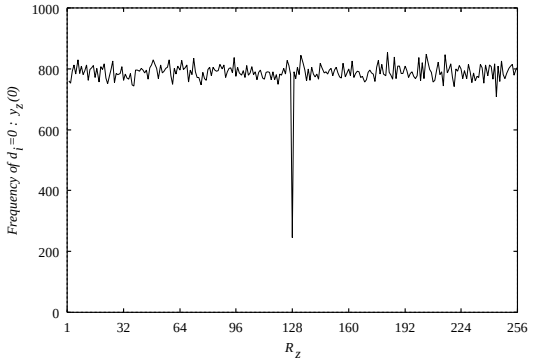
本手法では，透かし信号の埋込みに用いる周波数係数の組を秘密鍵として高い秘匿性を実現できる．

まず，Classic を用いて R_{128} を埋込み鍵とし，$M = 1024$，$p = 5$，$f = 5$ による埋込みを施した音楽データから透かしが正しく検出できるかどうか調べた．**図 13** (a) に，埋込みのない状態の Classic から得られる d_i の分布を示した．また，埋込みを施した状態の Classic から d_i の分布を調べると図 13 (b) が得られた．これらから図 13 (a) に比して図 13 (b) は明らかに d_i の分布のランダム性が増していることが分かる．そこで，R_1〜R_{256} によって得られた $d_i = 0$ の頻度すなわち $y_z(0)$ を抽出すると**図 14** のようになった．この図は，横軸に検出に用いた乱数鍵 R_z の種類（番号）を表し，縦軸にその鍵による検出値 $y_z(0)$ を示している．この結果から，埋込みに使用していない鍵による $y_z(0)$ の値は，約 800 程度になることが分かる．これは，埋込みを施していない場合の MDCT 係数の組から得られる差分値の分布が 0 に偏ることを示している．一方，正しい埋込み鍵を用いて透かしを検出すると $y_z(0)$ の値が約 250 程度まで低下していることが分かる．このように埋込み鍵の乱数列 R_{128} と検出スペクトル $y_z(0)$ の低下が一致することにより透かしの存在を証明できる．

また，3 章に述べたように本手法では複数の透かしを多重に埋め込むことができる．そこで，Classic に 2 種類の埋込み鍵 R_{64}，R_{192} として $M = 1024$，$p = 5$，$f = 5$ による透かしを多重に埋め込み，その音声データから，埋込み鍵 R_{64}，R_{192} を含む 256 種類の鍵



(a) Original



(b) Watermarked

図 14 透かし信号の検出
Fig. 14 Detection result of watermark.

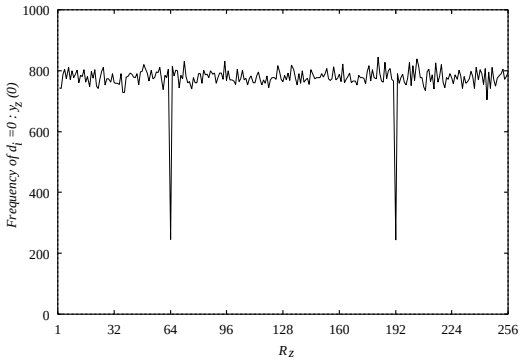


図 15 多重透かしの検出
Fig. 15 Detection result of multiplex watermarks.

R_1〜R_{256} による検出を行うと**図 15** が得られた．この結果から，透かしの埋込みに異なる乱数列を用いれば透かしを多重化ができることが分かる．

本手法によると埋込みに用いた乱数列が一致した場合に限り透かしを検出できる．したがって，埋込みに用いた乱数鍵を知らない第 3 者が透かしの存在を不正な手段により特定し，音質をほとんど劣化することなく透かし情報を除去することは難しい．

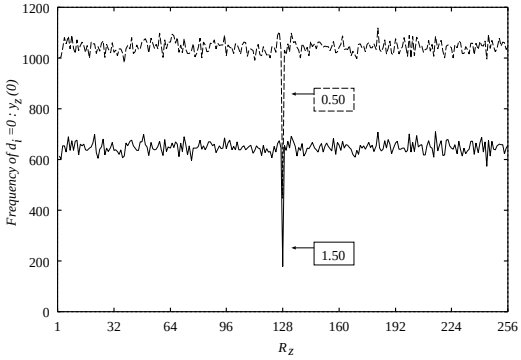


図 16 レベル変調による影響（$\gamma = 0.5$，1.5）
Fig. 16 Influence on watermark by amplification ($\gamma = 0.5$, 1.5).

4.5 レベル変調耐性

本手法によれば，MDCT 係数の組から得られる差分のランダム性が保持される程度のレベル変調であれば透かしが消失することはない．そこで，R_{128}，$M = 1024$，$p = 5$，$f = 5$ を用いて埋込みを施した Classic の波形スペクトルを定数 $\gamma = 0.5$，1.5 により増幅（減衰）し，透かしが消失しないかどうかを調べた．**図 16** の検出結果から，$\gamma = 0.5$ のときは周波数スペクトルの値が小さくなるため $y(0)$ の値が全般に増加し，グラフ全体が上にあがることが分かる．一方，$\gamma = 1.5$ のときは，周波数スペクトルの値が大きくなり $y(0)$ の値が減少するためグラフ全体が下にさがることが分かった．しかし，いずれの場合もグラフ全体が影響を受けるため，透かしの存在が判別できた．この結果からも本方式による電子透かしは，ある程度の音声波形の振幅レベルの変調に耐性を有することが分かる．

4.6 帯域通過フィルタ耐性

不特定の MDCT 係数に選択的に透かしを埋め込む利点として，帯域通過フィルタへの耐性が考えられる．本方式によれば制限帯域外に施された埋込み処理は，原理的にフィルタリングの影響を受けない．そこで，一般に利用されている離散フーリエ変換による帯域通過フィルタ[1)]を用いて再生音楽の帯域幅を 5 kHz から 17 kHz に制限し，透かし信号が消失しないかどうかを調べた．通常，このレベルの帯域制限を施すと音質の劣化を明らかに知覚できることに注意する．ここでは，$M = 1024$，$p = 5$，$f = 5$ として Classic に埋込みを施し，さらに帯域制限した音楽データからの透かし信号の検出結果を**図 17** に示した．この図から，本方式を原理とする電子透かしは帯域通過フィルタ処理に耐性を持つことが分かる．

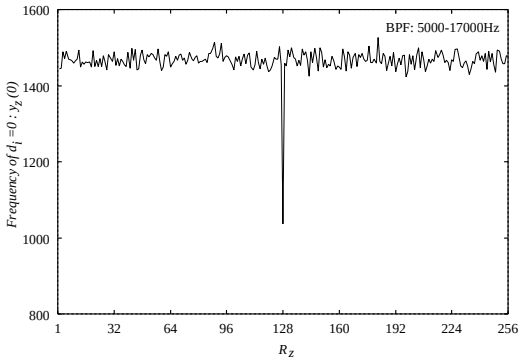


図 17 帯域通過フィルタによる影響（5〜17 kHz）
Fig. 17 Influence on watermark by bandpass filtering (5〜17 kHz).

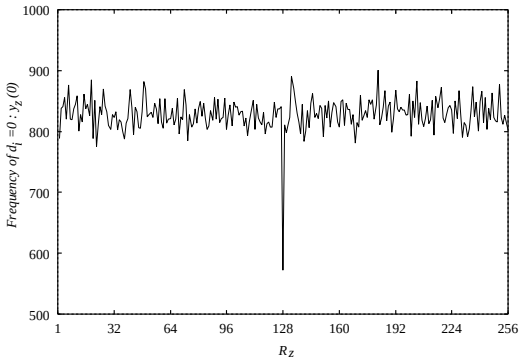


図 18 MP3 圧縮による影響
Fig. 18 Influence on watermark by MP3 compression.

4.7 高能率圧縮符号化の影響

周波数変換やサブバンド符号化などを原理とした高能率圧縮符号化方式である MPEG1 Audio LayerIII（MP3）のアルゴリズム[15]を用いて高度なデジタル信号処理への耐性を調べた．MP3 による符号圧縮は，インターネットを利用した高音質な音楽の配信ツールとして普及しており，その不正目的の利用について著作権保護上の問題が指摘されている．本実験では，インターネット環境において標準的に利用されている圧縮率が約 1/10 のビットレートの MP3 を用いた．**図 18** は，$M = 1024$，$p = 5$，$f = 5$ として埋込みを施した Classic を MP3 を用いて圧縮伸張し，その再生音楽から透かしを検出したものである．この結果から，MP3 圧縮符号化による冗長成分の除去によって透かしの検出率がやや低下するものの，正しい鍵による検出値 $y(0)$ が一般に低い値を示すことを確認できた．したがって，本方式は，ある程度の情報圧縮にも耐性を持つと考えられる．

4.8 Jitter Attack

音楽データに埋め込まれた電子透かしを破壊する攻撃として知られる Jitter Attack[13]によって，本手法による透かしが消失しないかどうかを調べた．Jitter Attack は，音楽の周波数帯域を不特定に 0.1%程度増減する攻撃法である．これは，人間の聴覚が微少なピッチの変動を知覚できない特徴を利用した攻撃法である．特に，直接拡散法やパッチワーク法などの透かし信号の検出を困難にするために有効である．まず，Classic に $M = 1024$，$p = 5$，$f = 5$ で透かしを埋め込み，Jitter Attack を施した再生音楽データを準備した．この音楽データから透かしを検出すると**図 19** が得られた．この結果から，提案方式のように周波数成分として透かし信号を埋め込むことにより，Jitter Attack などによるフレーム同期の崩れに対して耐性を得られることが分かる．ただし，音質劣化が知覚できるような大きなピッチ変動への耐性を確保するには，フレーム同期信号の挿入などによる対策が必要であると考える．

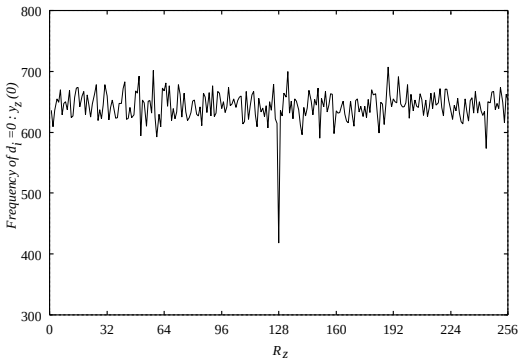


図 19 Jitter Attack による影響
Fig. 19 Influence on watermark by Jitter Attack.

5. おわりに

この論文では，CD などに用いられる高品質音声データに透かし情報を密かに埋め込む方法を提案した．本手法によれば，音質をほとんど損なうことなく，ある程度の符号圧縮やデジタル信号処理にも耐える電子透かしを埋め込むことができる．よって，不正コピーされた音楽データを調べて透かし情報を検出することにより不正行為の事実を特定できる．一方，利用者は，透かしの存在を知覚できないうえに不正な手段を用いてその存在を知ることも難しい．このような電子透かしが広く認知されれば，不正行為を心理的に抑止する効果も期待できる．

本提案方式による透かし検出処理には，各フレームごとの埋込みを施した周波数成分の位置（高さ）に関する情報があればよい．そのため，検出の際に透かし

信号の強さそのものを再現する必要がない特徴がある☆．すなわち，式 (10) の一例に示した埋込み制御系列 W_i の生成に関する自由度が大きい．これは，提案方式への埋込み強度の適応化処理の導入が容易であることを意味している．また，これまでに時系列標本値へ直接埋め込む電子透かし技術が数多く検討されている．本研究では，埋込み対象の周波数成分をランダムに直接制御しているが，あらかじめ準備した埋込み制御系列 W_i を時系列波形値として構成（逆 MDCT 処理）し，従来方式のように時系列へ直接埋め込むことも容易である．その場合でも，全体的な波形の形状（周波数成分）として埋め込む提案方式の方が，透かし信号を検出しやすいと考えられる．今後は，これらの特性を考慮した適応化処理などを用いて，提案方式に改良を加えていきたいと考えている．

参 考 文 献

1) 小澤一範：ディジタル移動通信のための高能率音声符号化技術，トリケップス (1992).
2) 松井甲子雄：電子透かしの基礎—マルチメディアのニュープロテクト技術，第 7 章，森北出版 (1998).
3) Boney, L., Tewfik, A.H. and Hamdy, K.N.: Digital watermarks for audio signals, *Proc. International Conference on Multimedia Computing and Systems*, pp.473–480 (1996).
4) 松井甲子雄，中村康弘，ナタウットサムパイブーン：音声通信への文字情報の埋め込み，第 18 回情報理論とその応用シンポジウム，pp.389–392 (1995).
5) 岩切宗利，松井甲子雄：適応差分 PCM 符号化における音声符号へのテキスト情報の埋込み，情報処理学会論文誌，Vol.38, No.10, pp.2053–2061 (1997).
6) 松井甲子雄，岩切宗利：低遅延符号励振線形予測符号化による音声符号への電子透かし，画像電子学会誌，Vol.27, No.5, pp.475–482 (1998).
7) 岩切宗利，松井甲子雄：共役構造代数符号励振線形予測による音声符号へのテキスト情報の埋込み，情報処理学会論文誌，Vol.39, No.9, pp.2623–2630 (1998).
8) 岩切宗利，松井甲子雄：スペクトル拡散と変形離散コサイン変換による高品質デジタル音声のための電子透かし法，情報処理学会論文誌，Vol.39, No.9, pp.2631–2637 (1998).
9) 富岡淳樹，中村高雄，小川　宏，高嶋洋一：マルチチャネルディジタルオーディオに対する電子透かし，1998 年電子情報通信学会情報・システムソサイエティ大会，D-14-4，p.323 (1998).
10) 松本　勉，井上大介，北林創太：演奏データファイル SMF への情報ハイディング方式，2000 年暗号と情報セキュリティシンポジウム，SCIS2000-C03 (2000).
11) 山内雪路：スペクトラム拡散通信，東京電機大学出版 (1994).
12) Bender, W., Gruhl, D. and Morimoto, N.: Techniques for data hiding, *Proc. SPIE*, Vol.202, pp.2420–2440 (1995).
13) Petitcolas, F.A.P., Anderson, R.J. and Kuhn, M.G.: Attacks on Copyright Marking Systems, *2nd Workshop on Information Hiding*, pp.218–238 (1998).
14) 筒井京弥：楽音・音声圧縮方式—ATRAC2，インタフェース，Vol.23, No.7, pp.134–142, CQ 出版 (1997).
15) Rao, K.R. and Hwang, J.J.: *Techniques and Standards for Image, Video, and Audio Coding*, Prentice Hall (1996). 安田　浩，藤原　洋（監訳）：デジタル放送・インターネットのための情報圧縮技術，共立出版 (1998).

(平成 12 年 8 月 4 日受付)
(平成 15 年 2 月 4 日採録)

岩切　宗利（正会員）

昭和 45 年生．平成 5 年防衛大学校情報工学科卒業．平成 10 年防衛大学校理工学研究科情報数理専攻修了．平成 11 年防衛大学校情報工学科助手．マルチメディアと情報セキュリティに関する研究に従事．博士（工学）．電子情報通信学会，日本音響学会，映像情報メディア学会，画像電子学会各会員．

松井甲子雄（正会員）

昭和 14 年生．昭和 36 年防衛大学校電気工学科卒業．昭和 40 年九州大学大学院工学研究科電子専攻修了．昭和 56 年防衛大学校電気工学科教授．平成元年同大情報工学科教授．この間暗号学，情報セキュリティ，電子透かし，音声・画像データの符号化に関する研究に従事．工学博士．電子情報通信学会，画像電子学会，映像情報メディア学会各会員．著書「電子透かしの基礎」（森北出版）で第 15 回電気通信普及財団賞受賞．

☆ 攻撃処理などにより周波数スペクトルが大域的な変調を受けても検出にあまり影響しない利点がある．