

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

岩 切 宗 利[†] 松 井 甲 子 雄[†]

この論文では、高品質デジタル音楽ソフトに電子透かしを埋め込む方法を提案する。まず、音声データをスペクトル拡散し、変形離散コサイン変換を行う。このとき窓関数に対応した信号成分のある高効率な特定周波数に注目し、透かし情報を埋め込むものである。この方法により透かし情報を埋め込むとその影響を逆変換により広帯域に拡散できるため、音声帯域内に生じるノイズの発生を著しく低減できる。しかも拡散に用いたPN系列が鍵として高い秘匿性を確保する役割を果たす。本手法を用いた実験では、60~70 dBの高音質を維持しつつ170ビット/秒程度の透かし情報を埋め込むことが可能であることを確かめている。

Digital Watermark Scheme for High Quality Audio Data by Spectrum Spreading and Modified Discrete Cosine Transform

MUNETOSHI IWAKIRI[†] and KINEO MATSUI[†]

This paper presents a digital watermark scheme for high quality audio data. It is constituted with spectrum-spreading by PN sequence and embedding of watermark by the modified discrete cosine transform. The watermark signal is uniformly spread over the whole window frame when it is converted to sound. Thus, the watermarked noise is suppressed under very low level. For example, the sound quality is attained more than 60 dB in the segmental SN ratio by our scheme. The watermark is also protected against attack because of private PN sequence. The embedded amount will be estimated about 170 bits per second in 44.1 kHz sampling rate.

1. はじめに

通常、音響データでは、サンプリング定理に基づいて音声を標本化し、量子化して線形パルス符号化(PCM: Pulse Code Modulation)¹⁾によりデジタル化している。特に、音楽ソフトは、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、非常に高い音質を保ったままデジタル化できる。

一方、このデジタルデータは容易に完全な形式で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となってくる。この対策として、人間が知覚できない形式で著作権情報を音楽ソフトに埋め込む試みがある²⁾。

たとえばBoneyら³⁾は、聴覚上のマスキング現象を利用した電子透かしの埋込み法を示し、また、松井ら⁴⁾は量子化雑音に見せかけて透かしを埋め込み、伝送する方法を提案している。しかし、これらの方法は音の振幅成分を直接利用して埋め込むため、ノイズをともしやすく高品質な音楽ソフトには不適切であるとともに秘匿性が問題となる。さらに、岩切ら⁵⁾は、国際標準規格G.726により圧縮された音声符号に電子透かしを埋め込む方法を提案しているが、この方法は電話級の低音質の場合に適用するものである。

そこで、この論文では高品質を要求される音楽ソフトに適切な方法として、周波数成分に著作権情報を埋め込むことを考える。

まず、電子透かしの秘匿性を確保するために各コンテンツごとに異なる疑似乱数系列を用いて直接スペクトル拡散⁶⁾を行う。そして、拡散された広帯域に及ぶ音声波を変形離散コサイン変換(MDCT: Modified Discrete Cosine Transform)⁷⁾により周波数表現する。このように展開した高品質音楽ソフトのある特定周波

[†] 防衛大学校情報工学科

Department of Computer Science, National Defense Academy

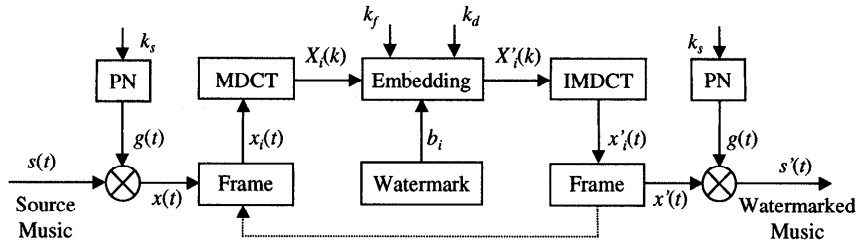


図1 透かしの埋込み処理
Fig. 1 Watermark embedding diagram.

数成分に透かし情報を埋め込む方法を提案する。この方法によると復元音声にフレーム歪みを生じることなく透かしビットを埋め込むことができる。さらに同一疑似乱数系列を用いた逆拡散によりこの透かし情報を広帯域に拡散するので、秘匿性を高めるとともにノイズを低く抑制することが可能となる。

まず、2章ではスペクトル拡散および変形離散コサイン変換を用いた透かしの埋込み方法を示す。次に、3章において埋込み処理による影響について考察する。また、4章では、この原理に基づいたシステムを構成し、2, 3の実験を行った結果を示す。

2. 透かしの埋込み方法

この論文で用いた電子透かしの処理の概要を図1に示す。まず、電子透かしの秘匿性を確実にするためスペクトル拡散の直接拡散 (DS: Direct Sequence) 方式⁶⁾を用いる。直接スペクトル拡散法には、疑似乱数 (PN: Pseudo-random Numbers) 系列が必要となるので、このPN系列の生成鍵をもって電子透かしの固有の秘密鍵とする。

サンプリング時刻 t における音声信号を $s(t)$ 、 ± 1 の値をランダムにとる2値のPN系列を $g(t)$ とする。また、 $g(t)$ を発生させるための鍵を k_s とする。このとき、 $g(t)$ による拡散信号 $x(t)$ は

$$x(t) = s(t) \cdot g(t) \quad (1)$$

である。この広帯域に拡散された音声信号 $x(t)$ を変形離散コサイン変換する。変形離散コサイン変換⁷⁾は、図2に示すように 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、フレームを互いに重複させてフレーム歪みを抑制するための処置である。

すなわち、 i 番目のフレームにおける MDCT 係数 $X_i(k)$ は

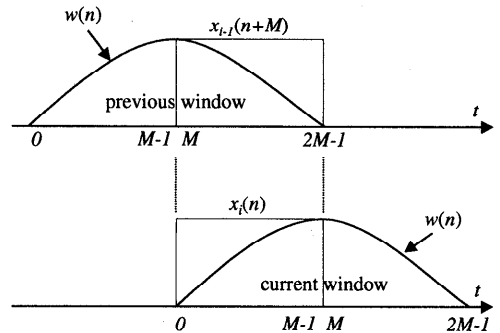


図2 フレーム間の信号干渉
Fig. 2 Signal interference with the previous frame.

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

$$0 \leq k \leq M-1, 0 \leq n \leq 2M-1$$

により求める。ただし、窓関数 $w(n)$ および MDCT 基底 $c(k, n)$ は、それぞれ

$$w(n) = \sin \left(\frac{\pi(2n+1)}{4M} \right) \quad (3)$$

$$0 \leq n \leq 2M-1$$

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

$$0 \leq k \leq M-1, 0 \leq n \leq 2M-1$$

である。ここで透かしビット b_i を埋め込むために埋込み用の周波数鍵 k_f ($k_f = 0, \dots, M-1$) およびスケール鍵 k_d ($k_d \neq 0$) を導入する。

まず、図3に従い $X_i(k_f)$ を k_d で量子化し整数値 e_i を求める。

$$e_i \leftarrow \lfloor X_i(k_f)/k_d \rfloor \quad (5)$$

このとき、もし e_i が奇数でかつ $b_i = 0$ 、または e_i が偶数でかつ $b_i = 1$ ならば

$$e_i \leftarrow \lceil X_i(k_f)/k_d \rceil \quad (6)$$

とする。この修正された e_i を用いて透かし情報を含

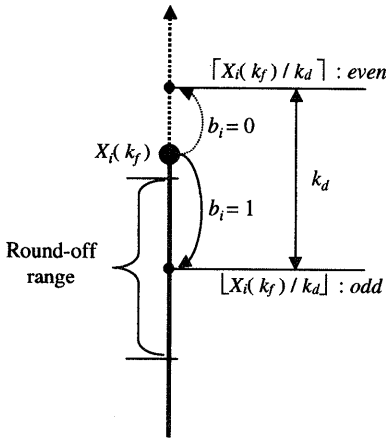


図3 埋込み規則
Fig. 3 Embedding rule.

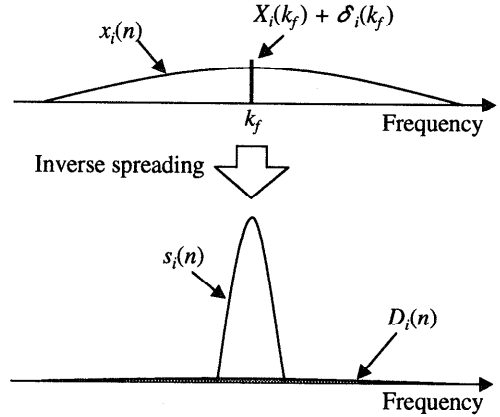


図4 透かしの影響
Fig. 4 Influence of watermark.

むフレーム周波数データ

$$X'_i(k_f) \leftarrow e_i \cdot k_d \quad (7)$$

を求める。このように埋込みを施した MDCT 係数 $X'_i(k)$, $k = 0, \dots, M - 1$ を順変換と同じ窓関数を用いて各フレームの波形要素 $x'_i(n)$ に変換する。すなわち、

$$x'_i(n) = w(n) \sum_{k=0}^{M-1} c(k, n) X'_i(k) \quad (8)$$

$$0 \leq k \leq M - 1, 0 \leq n \leq 2M - 1$$

これらの $x'_i(n)$ から、時系列信号 $x'(t)$

$$x'(n + iM) = x'_{i-1}(n + M) + x'_i(n) \quad (9)$$

$$0 \leq n \leq M - 1$$

を復元する。

このように、MDCT 係数 $X_i(k_f)$ に埋込み操作をすることは、図 4 のように狭帯域の雑音干渉 $\delta_i(k_f)$ を与えたことに相当する。よって、透かし(雑音)の帯域幅は音声復元時に M 倍に拡大され、透かしの雑音電力は非常に小さくとなると考えられる。

ところで、同じ PN 系列 $g(t)$ には、 $g^2(t) = 1$ の性質があるので、

$$x'(t) \cdot g(t) = s'(t) \cdot g^2(t) = s'(t) \quad (10)$$

を得ることができる。この操作を各フレームに対して繰り返し適用することにより、埋込み済みの出力波形が得られる。

このとき、透かしの埋込み量 B_r は、音声の標準化周波数 f_s を用いて、

$$B_r = \frac{f_s}{M} \text{ [bit/s]} \quad (11)$$

により求められる。ただし、これは各フレームに対して 1 ビットの埋込みを施した場合であることに注意する。よって、音質の劣化と引換えに、各フレームに対して複数ビットを埋め込むことは容易に可能である。このときの埋込み量は、各フレームごとの埋込みビット数に比例する。

一方、透かしビットの抽出は図 5 に示す手順で行われる。まず、透かしを含む音声信号 $s'(t)$ を埋込みに用いた PN 系列 $g(t)$ により拡散して DS 信号 $x'(t)$ を求める。次に、各フレームからその MDCT 係数 $X'_i(k_f)$ を抽出し、 k_d を用いて量子化する。すなわち、

$$e'_i \leftarrow \text{Round}(X'_i(k_f)/k_d) \quad (12)$$

ただし、 $\text{Round}(r)$ は、四捨五入法による整数値である。ここで e'_i が偶数か奇数かを調べ、偶数の場合は $bo_i = 0$ 、奇数の場合は $bo_i = 1$ として透かしビット bo_i を復元することができる。このとき、復号に用いる PN 系列が埋込みに用いた $g(t)$ と異なるならば、正しい MDCT 係数分布を得られないため、埋め込まれた透かし情報を正しく復元できないことに注意する。したがって、本方式では、埋込みに用いた $g(t)$ を抽出鍵として高い秘匿性を確保できる。

3. 埋込み処理の影響

ここで、透かしビットの埋込みによる影響について考察する。まず、MDCT 係数への埋込み操作により生じた量子化誤差を $\delta_i(k)$, $k = 0, \dots, M - 1$ とする。

$$\delta_i(k) = X'_i(k) - X_i(k) \quad (13)$$

このとき、各フレームの波形要素 $x_i(n)$ に生じる埋込みの影響 $D_i(n)$ は、

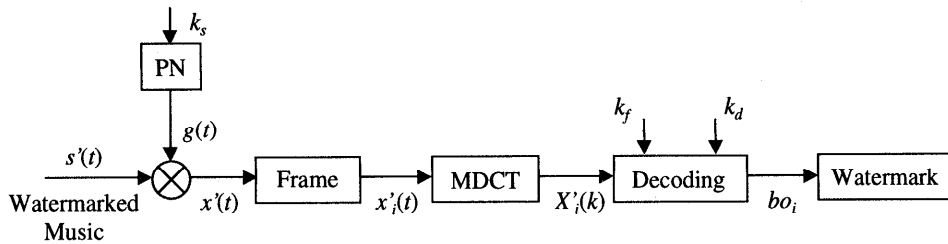


図5 透かしの抽出処理

Fig. 5 Watermark decoding diagram.

表1 実験音声
Table 1 Sound for experiment.

Name	Samples	Sec
Classic	44,0832	10
Jazz	44,0832	10
Dance	44,0832	10

$$\delta_i(k) = \begin{cases} \delta_i(k_f) & k = k_f \\ 0 & k \neq k_f \end{cases} \quad (14)$$

および式(9)より

$$D_i(n) = w(n)c(k_f, n)\delta_i(k_f) \quad (15)$$

$$0 \leq n \leq 2M - 1$$

となる。したがって、埋込みが復元信号に与える影響 $D(n)$ は

$$D(n) = D_{i-1}(n + M) + D_i(n) \quad (16)$$

$$0 \leq n \leq M - 1$$

である。また、 $\delta_i(k_f)$ は、 k_f に依存せず $|\delta_i(k_f)| \leq k_d$ になる。ここで、デジタル波形値に生じる誤差は、 $D(n)$ を整数化した値 $\hat{D}(n)$,

$$\hat{D}(n) = \text{Round}(D(n)), 0 \leq n \leq M - 1 \quad (17)$$

となる。

4. 実験結果と考察

音質の高さに価値がある音楽ソフトに埋込みを施す場合、埋込みによって音質が劣化しないことが重要である。そこで、実際に高品質の音楽データに埋込みを施し音質に与える影響を調べてみよう。

4.1 実験音声

表1に示した実験音声は、44.1 kHz、16 bit で量子化したものである。ただし、音楽ソフトは、通常ステレオ音であるので、その片側成分のみに埋込みを施している。

4.2 音質の評価法

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⁾。

$$\text{SNR} = 10 \log_{10} \left\{ \frac{\sum_m So^2(m)}{\sum_m Er^2(m)} \right\} \quad (18)$$

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 SNR_{seg} (Segmental SNR) を用いた¹⁾。

$$\text{SNR}_{\text{seg}} = \frac{1}{N_f} \sum_{f=1}^{N_f} \text{SNR}_f \quad [\text{dB}] \quad (19)$$

N_f は測定区間のフレーム数を表し、 SNR_f は、 f 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、 $\text{SNR}_f = \infty$ の音声フレームを除外して測定した。

4.3 実験結果

本実験では、 $M = 256$ としたため B_r は約 172 ビット/秒である。この埋込みを施した再生音声と埋込みのない再生音声の音質を比較検討する。

まず、 $k_d = 1$ に固定し、 k_f を変化させて Classic に埋込みを施した場合の SNR_{seg} を調べると図6の結果が得られた。この結果から、いかなる周波数成分 k_f を用いても同程度の高い音質を維持していることが分かる。

また、ほかの埋込み鍵でも同様の特徴が現れるかどうかを検討する。ここでは、図6と同様、Classic に対して埋込みを施した場合の SNR_{seg} の変化を図7および図8に示した。これらの結果から異なる鍵 (k_d, k_s) を用いても、すべての鍵 k_f に対して同程度の音質が得られていることが明らかである。したがって、本方式では埋込みを施す周波数成分 k_f に依存せず同程度の音質を得られることが分かる。

次に、 $k_d = 2$, $k_f = 128$, $k_s = 128$ として Classic に透かしの埋め込んだ音声データに対し、異なる k_s を用いて透かしの検出を試みた場合の検出率を図9に

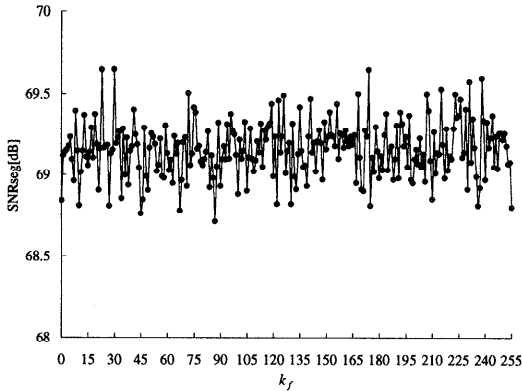


図6 k_f に対する SNR_{seg} ($k_d = 1, k_s = 1$)
Fig. 6 SNR_{seg} to k_f ($k_d = 1, k_s = 1$).

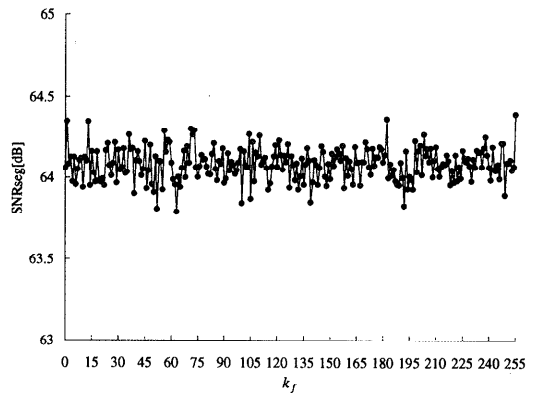


図8 k_f に対する SNR_{seg} ($k_d = 2, k_s = 2$)
Fig. 8 SNR_{seg} to k_f ($k_d = 2, k_s = 2$).

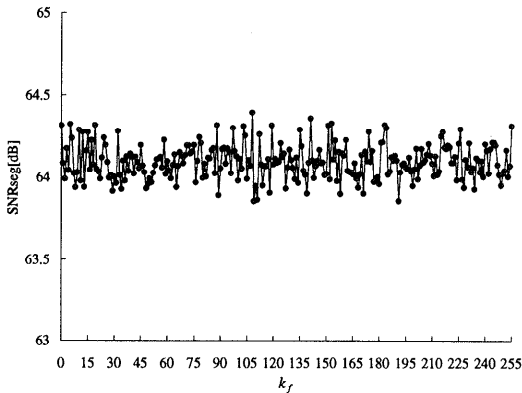


図7 k_f に対する SNR_{seg} ($k_d = 2, k_s = 1$)
Fig. 7 SNR_{seg} to k_f ($k_d = 2, k_s = 1$).

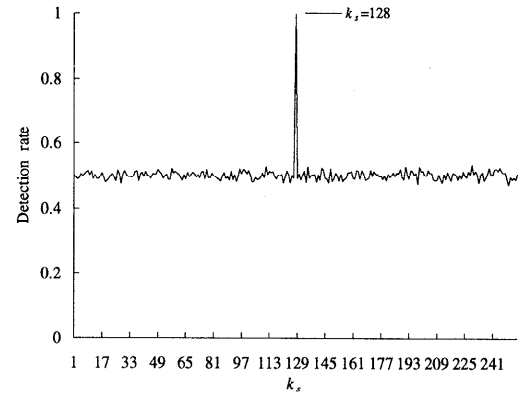


図9 透かしの検出率
Fig. 9 Detection rate of watermark.

示す。この検出率は、埋め込まれた透かしビットと抽出された透かしビットの一致率を意味する。図9の結果から当初の鍵 $k_s = 128$ で生成したPN系列で復号すると検出率1で完全に復号できるが、他の異なった鍵 k_s では検出率0.5となり、0と1がランダムに出現するので透かし情報を正しく抽出できないことが分かる。よって、拡散に用いた $g(t)$ を鍵として高い秘匿性を確保できると考えられる。しかし、不正者が $g(t)$ と相関の高い系列を用いた場合、透かし情報の存在を発見することも考えられる。したがって、 $g(t)$ には相関の高い系列を容易に生成できないようなPN系列を用いることが望ましい。

次に、 $k_f = 128$ として k_d を変化させた場合の音質を図10に示す。この図から k_d を増大させると復元音質 SNR_{seg} は低下することが分かる。したがって、高音質を維持するためには k_d を1~3程度の低い値にすることが望ましい。音楽ソフトの商品的価値を失

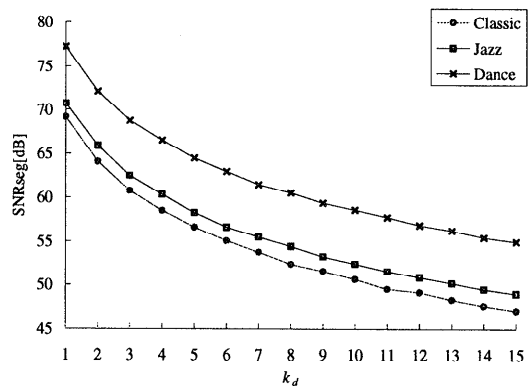
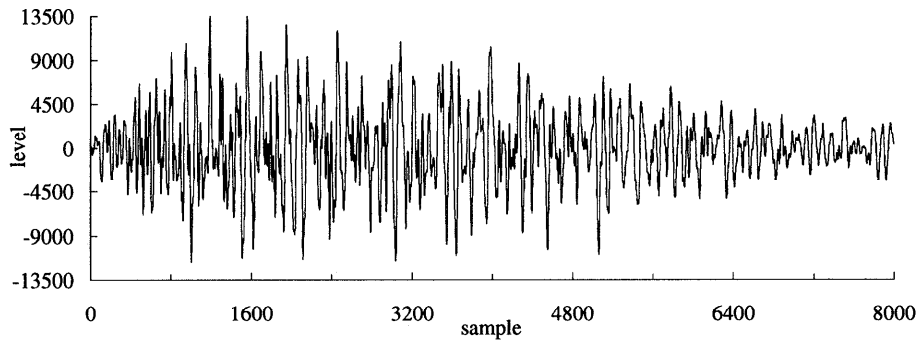
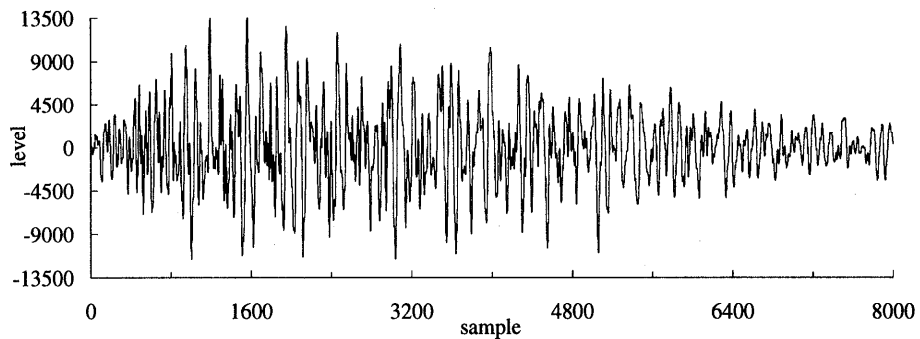


図10 k_d に対する SNR_{seg} ($k_f = 128, k_s = 1$)
Fig. 10 SNR_{seg} to k_d ($k_f = 128, k_s = 1$).

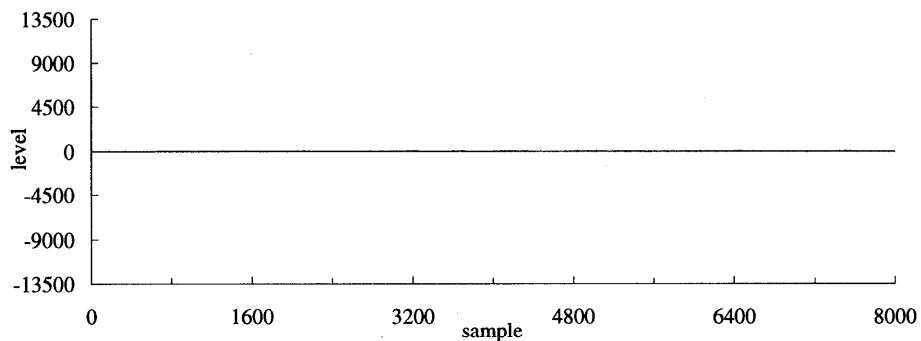
わずに(音質を劣化させずに)デジタルコピーした場合、 k_d を低い値としても透かし情報を取り出すことは可能である。また、音質劣化(損失圧縮)しても透かしを取り出す必要があるならば k_d を大きな値にす



(a) Decoded waveform of piano sound without watermark



(b) Decoded waveform of piano sound with watermark



(c) Difference waveform

図 11 音声波形の比較

Fig. 11 Comparison of sound waveform.

ればよい。しかし、音楽ソフト自体の商品的価値を低下させることになる。

さらに再生波形の一部を切り出して埋込み処理が波形の形状に与えた影響を観察してみる。図 11 に埋込みのない再生音声波形 (a) と埋込みを施した再生音声波形 (b) およびそれらの差分波形 (c) を示した。これら波形は、Jazz におけるピアノの演奏音に相当する部分である。この差分波形 (c) から (a) と (b) の波形にほとんど違いがないことが分かる。また、本方式を用いて通常公開される音楽ソフトは埋込みのあるもの

のみであることに注意する。よって、差分波形を通常的手段で得ることはできない。したがって、再生波形の形状から不正に埋込みのある音楽ソフトを特定することは難しいと考えられる。

5. むすび

この論文では、音楽ソフト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、音楽ソフトの音質を損なわずに高い秘匿性を持つ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ソフトを調べて透かし情報を検出することで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ゑに不正な手段でその存在を知ることも難しい。よつて、不正行為を心理的に抑止する効果も期待できる。

参 考 文 献

- 1) 小澤一範：デジタル移動通信のための高エネルギー音声符号化技術，トリケップス(1992)。
- 2) 松井甲子雄：デジタル透かし，画像電子学会誌，Vol.26, No.3, pp.266-274 (1997)。
- 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，ナタウツサムパイブーン：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め込み，情報処理学会論文誌，Vol.38, No.10, pp.2053-2061 (1997)。
- 6) 山内雪路：スペクトラム拡散通信，東京電機大学出版(1994)。

- 7) 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 (1997)。

(平成10年1月5日受付)

(平成10年7月3日採録)



岩切 宗利

昭和45年生。平成5年防衛大学校情報工学科卒業。平成10年同大理工学研究科情報数理専攻修了。



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。著書「画像深層暗号」(森北出版)。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。