

デジタル音楽への電子透かしの可視化法

岩 切 宗 利[†] 松 井 甲 子 雄[†]

本論文では、高品質デジタル音楽ソフトに埋め込まれた電子透かしを視覚情報として検出する手法を提案する。通常、音楽ソフトは、時系列上の波形値である PCM データとして流通している。したがって、従来の電子透かしでは、これを一次元軸上の値として署名を施したうえで配布し、不正に横流しされたデータから統計的に透かしの存在を検出する方法が一般的であった。そこで、この研究では音楽データを 2 次元メモリ上に配列したとき、得られる相互関係に着目し、画像情報として署名を施す音楽への電子透かし法を考える。この方法によると 1 次元音声から抽出した透かしデータを 2 次元の画像として視覚に訴えることができる。よって、再標本化や切り抜きなどの編集操作が施された場合や雑音が混入した場合でも、視覚によるパターン認識の優れた特性を生かし、電子透かしの認識率を向上できる利点がある。さらに、インターネットでの音楽配信に不可欠な MP3 などの高能率符号化による透かし信号の消失を補填できることも確認できた。本手法を用いた実験では、データ量を 128 kbit/s まで MP3 圧縮し、伝送・受信後に再生した。この音楽データから 256 × 256 画素の署名画像を抽出したとき、その透かし情報を十分に認識できることを確かめている。

Visualizing Technique of Digital Watermarks to Audio Data

MUNETOSHI IWAKIRI[†] and KINEO MATSUI[†]

In this paper, we propose how to embed a copyright stamp image into audio data and its visualizing technique of watermark. When an analog audio signal is sampled and quantized, the output codes from the quantizer are blocked by the $M \times N$ samples and the specified bit-plane of the image is controlled by watermark. The multiplexed signals sound well for us and thus anyone is not aware that the watermark is being simultaneously existed in the audio data. The visual detecting scheme brings us a robust watermark for copyright protection, thus the watermark is proof against such attacks as re-sampling, clipping, noisy channel and lossy compression. This visualizing method would be also available for digital audio watermarking.

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も埋

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の埋め込み伝送する方法を提案している。さらに、岩切ら^{5)~8)}は高能率な圧縮符号化に関する国際標準規格による音声符号や音楽ソフトなどの高品質な音楽データに適した電子透かし法を提案している。

これらによれば、聴感的な音質の劣化をほとんど感じさせることなく、大量の情報を埋め込むことが可能である。しかし、著作権侵害の有無を調べるには、復号したビット列のパターンを照合し、統計的に判定する必要がある。したがって、これらの実用化には、復号ビット列の誤り訂正に関する検討が不可欠であった。

そこで、本論文では 1 次元の音楽データ系列に埋め込まれた署名を冗長さの大きい 2 次元画像情報として確認する方法を提案し、不正行為を心理的に抑止することを考える。この方法によれば、透かしの検証に視

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

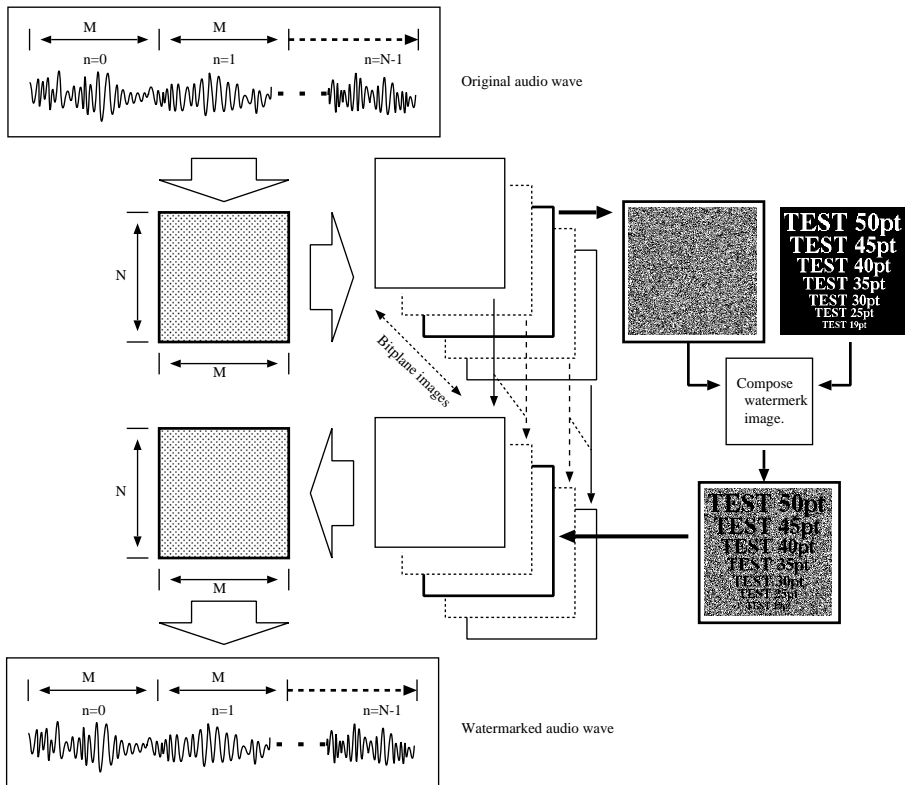


図 1 透かしの埋込み処理

Fig. 1 Watermark embedding diagram.

覚パターン認識の優れた特性をうまく活用できるため、透かしの強度を構造的に高めることが可能である。

まず、著作権情報を示す 2 値画像を準備し、その画素を 1 次元データとして走査する。そのビット情報に従ってデジタル化された音楽データの指定ビットプレーンのみを制御する。この操作を音楽データ全体に対して反復して行い埋込みを完了する。復号に際し、1 次元音楽データに埋め込まれた透かし情報を指定ビットプレーンから復号し、画像として 2 次元メモリ上に再配列する。このメモリに貯えられた著作権情報を画面に表示すれば、局所的にビット誤りなどが生じても画像全体として著作権情報の存在を視覚で十分に判断することができる。これは、人間の視覚によるパターン認識の特徴を利用したものであり、今後のデジタル音楽に対する電子透かしの検出法として有用であると考える。

まず、2 章では署名画像をデジタル音声データに埋め込む方法および埋込みによる影響について示す。次に、3 章において埋め込まれた画像を復号表示する手順を示す。また、4 章では、この手順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. 透かしの埋込み法

2.1 埋込みの原理

本論文で用いた電子透かしシステムの概要を図 1 に示す。まず、著作権を明示するために、大きさ $M \times N$ 画素の 2 値画像を準備する。この画像の座標 (m, n) の画素値を

$$b(m, n) = \begin{cases} 0, & \text{if black pixel} \\ 1, & \text{if white pixel} \end{cases} \quad (1)$$

$$0 \leq m \leq M-1, 0 \leq n \leq N-1$$

とする。ここで、 k ビットに量子化された音声標本値を $s(t)$ とし、その第 i ビット目、すなわち 2^i の重みを持つビット値を $s_i(t)$ とする。ただし、 $0 \leq i \leq k-1$ である。そこで、 $b(m, n)$ の値により

$$s'_i(t) = \begin{cases} s_i(t), & \text{if } b(m, n) = 0 \\ 0, & \text{if } b(m, n) = 1 \end{cases} \quad (2)$$

のように $s_i(t)$ の値を制御して署名済み音声データ $s'(t)$ を合成する。ただし、議論を簡単にするために標本時刻 t は、座標 (m, n) と

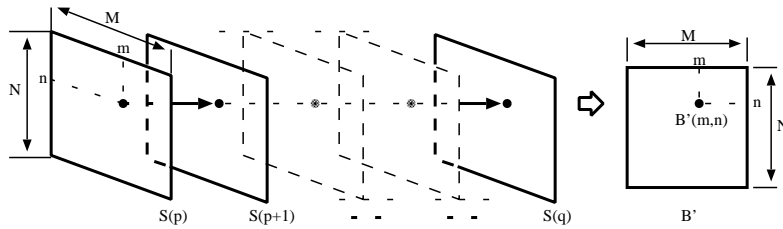


図2 透かし情報の加算過程

Fig. 2 Summing-up of watermarked pixels in series of frames.

$$t = n \cdot M + m \quad (3)$$

$$0 \leq m \leq M-1, 0 \leq n \leq N-1$$

の関係を満たす整数とする。

2.2 埋込みの影響

ここで述べた手法による署名済み音声 $s'(t)$ と原音声 $s(t)$ の差分値 $\delta_i(t)$ は,

$$\delta_i(t) = \begin{cases} 0, & \text{if } b(m, n) = 0 \\ 0, & \text{if } b(m, n) = 1 \text{ and } s_i(t) = 0 \\ \pm 2^i, & \text{if } b(m, n) = 1 \text{ and } s_i(t) = 1 \end{cases} \quad (4)$$

となる。また、著作権画像全体を1回埋め込むために必要なサンプル数 $M \cdot N$ の音声区間を透かしフレームとし、時系列順に $S(0), S(1), S(2), \dots$ とする。

ここで、 j 番目の透かしフレーム $S(j)$ に埋込みを施すとき、 $b(m, n) = 1$ かつ $s_i(t) = 1$ である音声標本数を $G_1(j)$ とする。このとき、埋込み操作によって影響を受ける標本値の含有率 $P(j)$ は、

$$P(j) = \frac{G_1(j)}{M \cdot N} \quad (5)$$

である。したがって、 $S(j)$ が埋込み操作によって受ける量子化誤差の平均二乗値 $\Delta(j)$ は、

$$\Delta(j) = 2^{2i} \cdot P(j) \quad (6)$$

である。よって、各標本値 $s(t)$ に対する埋込みビット位置が下位であるほど埋込みによる音質の低下は少ないことが分かる。しかし、埋込み位置が下位ビットであるほど、一般に各種信号変換の影響を受けやすい。したがって、埋込みを施すビット位置は、音質と透かしの強度の観点から適切に決定することが望ましい。

3. 透かしの復号法

3.1 視覚的復号法

ここでは、埋込みを施した音声データから透かし情報を検出する方法について述べる。まず、署名済み音声データ $s'(t)$ から 2^i の重みを持つビット値 $s'_i(t)$ を復号し、

$$b'(m, n) = s'_i(t) \quad (7)$$

とする。ただし、

$$\begin{cases} m = t - M \cdot n \\ n = \lfloor \frac{t}{M} \rfloor \end{cases} \quad (8)$$

である。この方法に従って得られた $b'(m, n)$ を $M \times N$ の2次元メモリに格納する。1フレーム分の出力を得たとき、それを2値画像として表示する。これにより1次元音楽データとして不可視であった透かし信号を画像パターンとして確認できる。

したがって、ここで述べた復号法によれば、人間のパターン認識能力により、編集操作やある程度のデータ誤りが生じた場合でも問題なく著作権情報の存在を確認できると考えられる。しかし、ここでは時系列上の標本値に埋込みを施しているため、データ圧縮などによって波形が大きく変動すると透かしの検出は困難になると考えられる。そこで、抽出した著作権情報を積分処理し、容易に視覚で検出できるように改善する。

3.2 透かし情報の積分処理

まず、連続した透かしフレームの系列 $s(j), j = 0, 1, \dots$ から、ある部分系列を取り出し、 $\{s(j) \mid j = p, p+1, \dots, q\}$ とする。この状態を図2に示す。ここに、 $0 \leq p < q$ とする。 $S(j)$ の任意の座標 (m, n) の画素値 $b'_j(m, n)$ は、0または1を表す2値データであり、これを j について p から q まで総和をとった値を $B'(m, n)$ で表す。すなわち、

$$B'(m, n) = \sum_{j=p}^q b'_j(m, n) \quad (9)$$

である。この $B'(m, n)$ を (m, n) 座標とする合成画像 B' (図2参照) で表す。一般に透かしフレーム $S(j)$ において署名部に該当する領域では $b'_j(m, n) = 0$ となり、背景部分のビット値は0または1のいずれかをとる。したがって、 $B'(m, n)$ は署名部に対応する領域と背景部に対応する領域では、 $p \sim q$ 間のフレーム数に比例した差が生じてくることになる。この処理から合成画像 B' は編集や圧縮、あるいはビット誤り、

透かしの削除攻撃などにより音楽データから部分的に消失した署名情報をフレーム間で補完し合い、全体として透かしの検出を容易にする効果が得られる。このとき、 $B'(m, n)$ は次の条件を満たしている。

$$0 \leq B'(m, n) \leq q - p + 1 \tag{10}$$

したがって、 B' を多値画像として表示する際には使用機器の画素表現範囲の上限に留意し必要ならば正規化しなければならないことに注意する。

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電子透かしを埋め込む場合、その操作によって品質を極端に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た。

また、署名済み音声データに対する編集操作や圧縮に関する2, 3の実験を行い、その強度に関する検討を実施した。

4.1 実験音声および署名画像

表1に示した実験音声は、44.1 kHzで標準化し、各標本値を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実験した。また、署名画像として図3を用いた。この画像は、256 × 256画素の2値画像であり、大きさの異なる7種類の文字パターンを例示した。

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

表1 実験音声

Table 1 Sound for experiments.

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



図3 実験用署名画像

Fig. 3 Experimental watermarks.

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⁾。

$$SNR = 10 \log_{10} \left\{ \frac{\sum_m So^2(m)}{\sum_m Er^2(m)} \right\} \tag{11}$$

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⁾。

$$SNR_{seg} = \frac{1}{N_f} \sum_{f=1}^{N_f} SNR_f \text{ [dB]} \tag{12}$$

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、 $SNR_f = \infty$ の音声フレームを除外して測定した。

4.3 実験結果

4.3.1 音声波形への影響

本実験では、 $(M, N) = (256, 256)$ としたため、透かしフレーム長は65536サンプルである。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、先に述べた埋込み手順に従って各ビットプレーン i に対して署名を施した音声の SNR_{seg} を調べると図4の結果が得られた。この結果から、原音声によって程度の差があるものの、一般に下位のビットプレーンほど埋込みによる影響は少ないことが明らかである。

さらに音声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する。図5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、図4に示した署名済みの音楽データ “Classic”

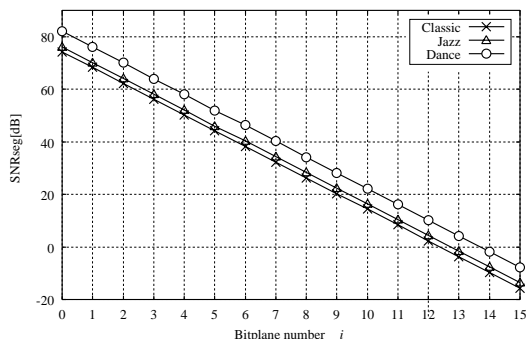


図4 埋込みビットプレーン i に対する SNR_{seg}

Fig. 4 SNR_{seg} to each bit-plane i for embedding.

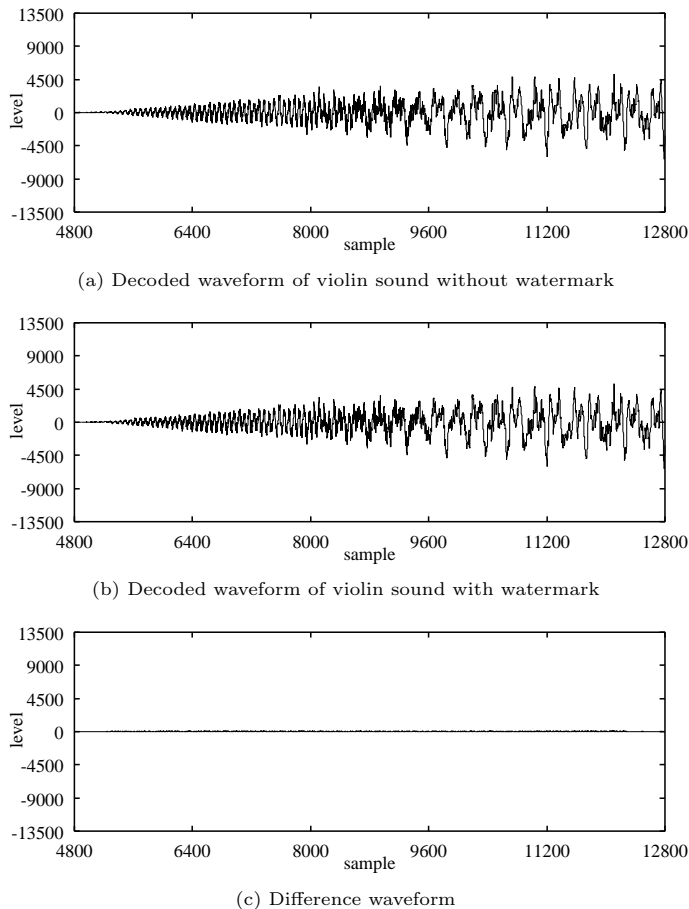


図5 音声波形の比較

Fig. 5 Comparison of sound waveform.

におけるバイオリン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を注意する。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。

4.3.2 透かしの強度

人間の視覚は、輪郭情報や構造を示す状態であれば十分に形状を認識できるという優れた能力を持っている。ここでは、視覚による透かしパターン検出の有効性について検討する。

まず、音楽データ “Dance” の $i = 0$ すなわち最下位のビットプレーンに対して埋込みを施し、単一の透かしフレームから署名情報を抽出した結果を図 6 に示す。この結果から埋め込まれた著作権情報を明確に認識できることが分かる。ただし、上位ビットに対す

る埋込みでも同様の結果が得られることに注意する。

さらに、この署名済みの音声データに対して約 20% の大量なデータ誤りを発生させ、その音楽データから透かしの抽出すると図 7 に示す透かし画像が得られた。この抽出画像から、音楽データに誤りが生じた局所的な情報を認識することはできない。しかし、透かしとしては、十分にその情報を保持していることが明らかである。

次に、音声データのブロック抽出および再標本化による影響について示す。まず、音楽データの冒頭部分を削除した状態のデータから透かしの復号すると図 8 が得られた。この結果から、透かし情報の位置がずれているものの、署名の内容を確認することは十分可能であることが分かる。また、22 kHz に再標本化した音声データから抽出した透かし情報を図 9 に示した。この結果から音楽データの標本数の減少にともなって、透かし画像も縮小された状態になることが分かる。しかし、透かしとして、その存在と記述内容を明確にす

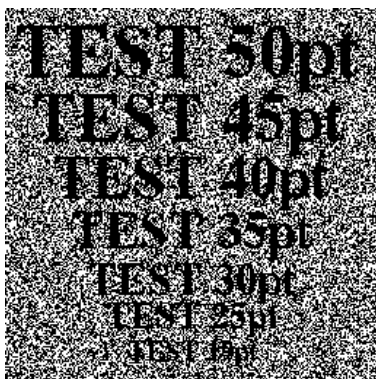


図6 最下位ビットからの署名画像
Fig. 6 Watermarks decoded from LSB.

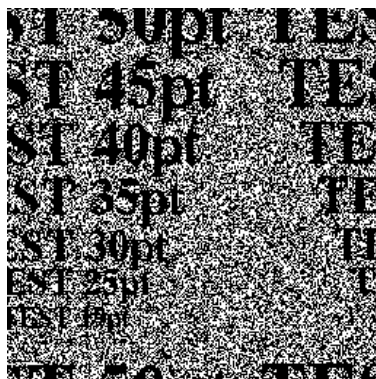


図8 切り抜き音声からの署名画像
Fig. 8 Watermarks decoded from clipped sound data.

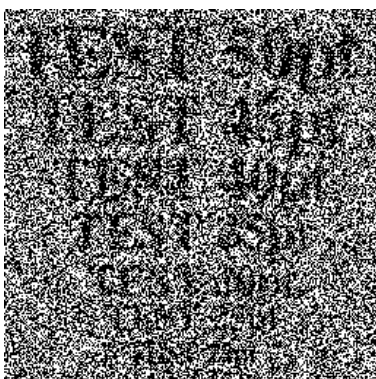


図7 透かしへのエラーの影響
Fig. 7 Error effect on watermark.

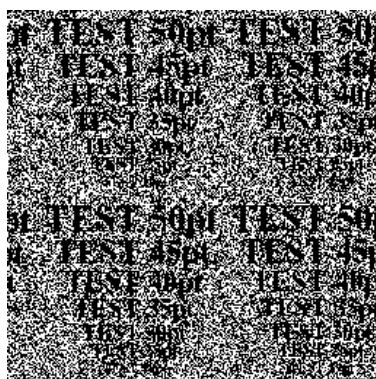


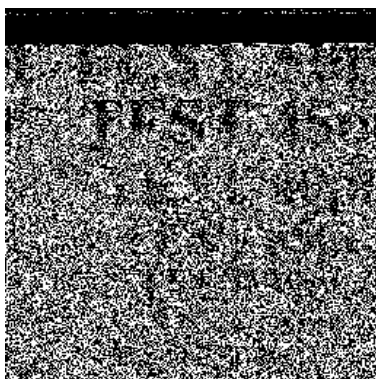
図9 再標本化による影響
Fig. 9 Influence of re-sampling.

ることは十分に可能である。

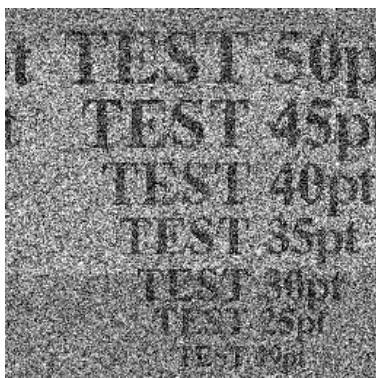
また、符号圧縮による影響について検討する。ここでは、近年 MP3 として脚光を浴びている MPEG-1 Audio layer3 を用いて実験した。まず、音楽データ “Classic” の下位から 7 ビット目に埋込みを施し、その音楽データを 128 kbit/s まで MP3 圧縮した。次に、MP3 圧縮符号から音声波形を再生し、透かしの抽出した。その結果、図 10 の署名画像が得られた。図 10 (a) は、これまでの実験と同様、単一の透かしフレームから復号した場合のものである。この結果では MP3 圧縮により署名情報が良好に認識できない状態になる。そこで、複数の透かしフレームから復号した署名情報を積分処理すると図 10 (b) が得られる。これは、埋込みに用いた原画像に比べてやや鮮明さに欠けるものの、署名の内容を十分に認識できる状態である。したがって、通常の復号方法では知覚できない透かし情報が、積分処理によって認識できる状態まで改善されることが分かる。よって、積分処理による復号は、高エネルギー圧縮などの波形が大きく変動する処理に対して有効であるといえる。

さらに、MP3 によって圧縮処理された場合における、埋込みを施すビット位置と透かし情報の認識率の関係について検討する。一般的には、埋込みビット位置が高次ビットであるほど透かしの強度は高いと考えられる。ここでは、埋込みビット位置の異なる音楽データをそれぞれ MP3 圧縮し、その再生データから署名情報を認識できるかどうかを判定した。その結果、表 2 が得られた。ただし、透かし画像のデータ量と実験音声のデータ量の関係から約 6.7 フレームを透かしの積分処理に用いている。この結果から、透かし画像を認識できる埋込みビット位置の下限は、実験音声によって異なることが分かった。これは、圧縮する音声データに適應して、MP3 の聴感的マスキングの特性が変化するためであると考えられる。これを解消する方策として、埋込み対象である音声データに応じて埋込みビット位置を適應化することが考えられる。

また、埋込みビット位置を透かしの存在を認識可能な最も低位のビットに固定した状態で、認識に十分な輝度差を得るために必要なフレーム数を調べたところ、いずれの音声についても 4 フレーム以上のフレーム数



(a) Watermarks extracted from a single frame



(b) Watermarks extracted from multiple frames

図 10 MP3 による圧縮の影響

Fig. 10 Influence of data compression by MP3.

表 2 MP3 符号化における透かしの再現可能なビットプレーン

Table 2 Visual bit-plane of watermarks under MP3 encoding.

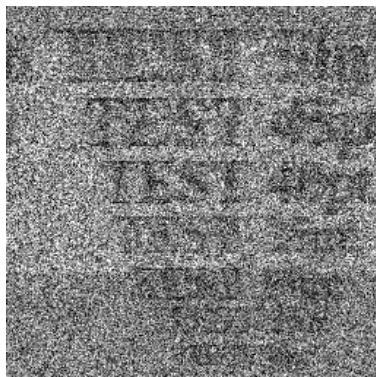
Bit plane i	4	5	6	7	8	9
Classic	x	x				
Jazz	x					
Dance	x	x	x	x	x	

が必要であった。よって、埋込みに用いる透かし画像のデータ量は、ある程度少ない方が望ましいといえる。ただし、署名情報を読み取るためには、ある程度の大きさの署名が必要であることに注意する。このことは図 10 (b) に示した透かし画像からも明らかである。

次に、オリジナル音声の周波数帯域を不特定に約 0.1%増減する Jitter Attack⁹⁾とよばれる変換処理について検討する。この Jitter Attack は、人間の聴覚が微小なピッチの変動を知覚できない特徴を利用したものである。まず、音楽データ“Classic”の下位から 7 ビット目に埋込みを施した音声データを準備し、それに対して Jitter Attack を行った。この変換された署名済み音声データから透かし画像を抽出すると図 11 (a)



(a) Influence on watermark by Jitter Attack



(b) Influence on watermark by MP3 compression and Jitter Attack

図 11 Jitter Attack による圧縮の影響

Fig. 11 Influence on watermark by data compression and Jitter Attack.

が得られた。この結果から Jitter Attack を施しても透かし情報を十分に保持できることが分かる。

さらに、MP3 圧縮と Jitter Attack による多重変換を施した署名済み音楽データから透かしの抽出を試みた。その結果、図 11 (b) が得られた。ただし、この抽出には積分処理を適用している。この透かし画像からその存在は確認できるものの、図 10 (b) と比較すると輪郭部分が不鮮明に見える。これは、Jitter Attack によって透かしのフレーム間同期が微少かつランダムに変動し、積分処理による隣接部分の輝度差が減少したためである。

5. む す び

この論文では、音楽ソフトに埋め込まれた 1 次元の署名情報を 2 次元画像として表示し、視覚で判別する一手法を提案した。本手法を音声への電子透かしの検出技術として適用すれば、圧縮や編集などの変換処理が施された音声データからでも透かし情報を容易に認識できる。よって、インターネットなどで不正に横流

しされた音楽データから著作権情報を検出し、違法行為の事実を特定するための一助として有用であると考える。

また、ここで用いたビットプレーンに対する埋込み手法は、視覚パターン認識による透かし検出法の有用性やその原理に関する議論を簡単にするためのものである。よって、埋込み手法については、音質に与える影響や秘匿性などの観点から、さらなる検討が必要である。すなわち、本手法の実用化に際しては、音質に与える影響や秘匿性などの観点から複数の埋込み手法を選定し、複合的に適用することが望ましいと考える。

このような電子透かし技術の存在が広く認知されれば、将来、著作権者に損害を与えるような違法行為を抑止する効果を期待できるであろう。

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：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) 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め込み，*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) 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)。

(平成 11 年 8 月 12 日受付)

(平成 12 年 3 月 2 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。日本音響学会会員。



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。