

可聴帯域を考慮した音楽ソフトへの可変型電子透かし

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の可聴帯域に除去可能な音声透かしを、また 20 [Hz] 以下の低周波帯域に秘匿透かしを埋め込む手法を提案する。インターネットを活用した音楽コンテンツのオンライン配信は、市場を拡大する手段の 1 つとして期待されている。本研究では除去可能な音声透かしを音楽データに埋め込むことを考える。この方法によれば、透かしメッセージを混入させたままのノイズな試聴用音楽ソフトをオンラインで配信できる。このコンテンツに対してユーザの要求により対価の支払いと交換に透かしメッセージを不可聴に抑圧し、同時に知覚できない秘匿透かし信号を低周波帯域に埋め込むことが可能である。本方式により再配布するわずらわしさもなく、高品質な音楽データを利用者に提供できる。一方、音声透かしを除去したコンテンツを不正に横流された場合には、不可聴な透かしが検証情報として利用できる仕組みになっている。本手法を用いてオンライン用透かしシステムを試作し、MP3 による高効率圧縮方式のもとで実験した。それらの結果を考察し、本手法の特徴を述べる。

Variable Watermarking Scheme to Digital Music by Classifying Audible Band

MUNETOSHI IWAKIRI[†] and KINEO MATSUI[†]

In this paper, we propose a watermarking technique for digital audio data to embed both audible and inaudible messages. The former is a variable watermark, and the latter is a secret watermark that consists of 20 or less Hz. Some customers wish to take a chance of hearing a music in advance on the network when buying it. For this purpose, we embed two kinds of watermarks which prevent malicious users from copying and distributing the contents illegally. The variable watermark is necessary to present a sample version and is deleted from the sample version via authorized procedures under contract. The secret watermark remains permanently in the contents but it would not be found except authorized key. Under this circumstances we can save time and money to re-distribute the contents for customers. It is shown that the watermarks stand against lossy MP3 compression algorithm on our experiments.

1. はじめに

急速なコンピュータ技術の発展とインターネット環境の整備は高品質なデジタルコンテンツの利用を容易にした。特にデジタル音声は身近なものとして広く普及している。また、近年のインターネットや大容量メディアによるデジタルコンテンツの配布は、その流通を高速かつ広範囲なものにし新たな市場の開拓に大きな役割を果たしている。しかし、デジタルデータはその再現性の高さからまったく劣化のない複製が容易に可能である。そこで近年、デジタルコンテンツの著作権保護上の問題が指摘され、その解決策の 1 つとして電子透かしが注目されている¹⁾。これは、流通コンテ

ンツに人間が知覚できない著作権情報を埋め込み、著作物の不正な横流しの特定に役立てるものである。

しかし、現在までに提案された電子透かしの大半は、画像や映像などの視覚的なコンテンツを対象としたものである²⁾。すなわち、デジタル音声を対象とする電子透かしの報告例は少なく、さらなる検討が必要とされている。

デジタル音声への電子透かしとして、松井ら³⁾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。また、Boney ら⁴⁾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、岩切ら⁵⁾は透かし信号をスペクトル拡散することで高音質かつ秘匿性の高い電子透かし法を実現している。ほかにも、反響音や音源定位を制御する手法^{6),7)}、圧縮音声符号に直接埋め込む手法^{8)~10)}などが検討されている。しかし、これらの手法は、いずれも透かしが

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

知覚されないことを主眼に検討されたものである。

ここで視点を転じ音楽ソフトの利用者の立場で議論する．音楽メディアの販売店では，利用者に商品を試聴する機会を提供している．それにより，利用者は商品の良さを知り，その価値に納得したうえで購入できる．同様に，インターネットを利用した電子配信にもこのシステムを導入する考えがある．しかし，そのままでは試聴用に提供された音楽データを継続利用される可能性が高く，不正に再配布されるおそれもある．この対策として，高品質な音楽ソフトの一部のみを試聴データとして提供することが考えられる．ただし，この方法では音楽ソフトの価値を十分に知らしめることができず購買意欲を高揚できないであろう．

そこで，本論文では可聴帯域の音声透かしを除去可能な可変形式で埋め込む手法を提案する．これは，音声透かしが聴こえる状態のまま試聴用音楽ソフトをオンラインで提供し，その購入契約により初めて音楽ソフトからその透かし信号を除去するものである．

まず，試聴用音楽ソフトに埋め込む可聴な透かし音声データを作成する．これを音楽の可聴帯域に対して反復的に埋め込み試聴用音楽ソフトを作成する．一方，透かし音声の周波数スペクトルを微妙に減衰した相殺成分と不可聴帯域の秘匿透かしを合成して鍵とする．これを試聴用音楽ソフトの音楽データに埋め込めば音声透かしの成分のみが抑圧される．すなわち，透かし音声のほとんどが鍵の埋め込みによって相殺され，試聴用音楽ソフトが高品質な音楽ソフトへ変換される．一方，試聴データに相殺成分が存在しない秘匿透かしの成分は，音声透かしの除去処理によって密かに埋め込まれる．本手法を用いれば，MPEG1 Layer3 (MP3) による高能率圧縮¹¹⁾を施しても透かし情報を保持することができる．

2章では音楽データに透かし音声の埋め込む原理およびその除去鍵を作成する手順について示す．3章では，透かしの除去時に非知覚状態の透かし信号を埋め込む方法を述べる．4章では，密かに埋め込まれた透かし情報の検出法を示す．5章では，本手法に基づいたシステムを用いて，2, 3の実験を行った結果について検討する．

2. 透かしの埋め込み法

2.1 透かし音声の埋め込み原理

この研究で用いた埋め込みシステムの概要を図1に示す．ここでは周波数分析に変形離散コサイン変換 (MDCT) を用いた¹²⁾．MDCTは，MPEGなどの高能率圧縮符号化技術に広く採用されている変換手法で

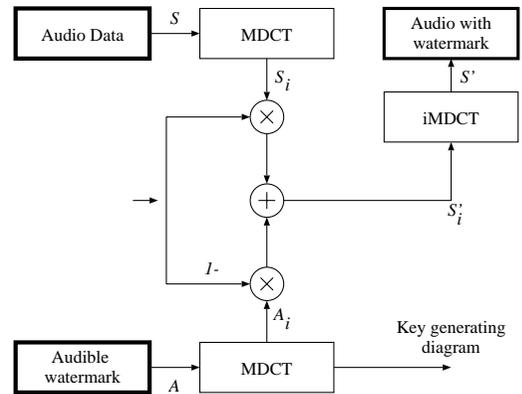


図1 埋め込み処理ブロック

Fig. 1 Watermark embedding diagram.

ある¹¹⁾．MDCTは， 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．音楽データ S の時系列標本値を $s(t)$ とする．このデータを M 個ずつまとめて1フレームとする． i 番目のフレームのMDCT係数 $S_i(k)$ を

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

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

で表す．ただし，窓関数 $w(n)$ およびMDCT基底 $c(k, n)$ は，それぞれ

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

$$c(k, n) = \cos\left(\frac{\pi(2k+1)(2n+M+1)}{4M}\right) \quad (3)$$

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

である．また，著作権を明示する透かし音声データを A とする．たとえば A として「このコンテンツはサンプルです」という音声などをデータ S に混入させることになる．この A の時系列標本値を $a(t)$ で表し， i 番目のフレームのMDCT係数 $A_i(k)$ を求める．このとき，音楽への可変透かしの埋め込みは，

$$S'_i(k) = \alpha \cdot S_i(k) + (1 - \alpha) \cdot A_i(k) \quad (4)$$

$$0 \leq k \leq M-1, 0 \leq \alpha \leq 1$$

によって施される．この式は，音楽ソフトの音波と透かし音声の周波数帯域での合成に相当する．また，制御係数 α を用いて音楽と音声透かし周波数成分比の和を1とし，再生波形値のオーバーフローや周波数スペクトルの異常な歪みを防ぐ工夫を施した．

一方，配信用音楽データ S' を再生するには， $S'_i(k), k = 0, \dots, M-1$ を順変換と同じ窓関数を用

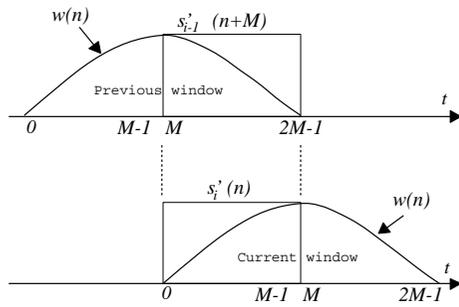


図2 フレーム間の信号干渉

Fig. 2 Signal interference with the previous frame.

いて

$$s'_i(n) = w(n) \sum_{k=0}^{M-1} c(k, n) \cdot S'_i(k) \quad (5)$$

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

により、各フレームの波形要素 $s'_i(n)$ へ変換する。この $s'_i(n)$ を図2に示すように

$$S'(n+iM) = s'_{i-1}(n+M) + s'_i(n) \quad (6)$$

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

として互いに干渉させることで、埋込み済み時系列信号 $S'(t)$ が復元される。こうして得られた時系列データ $S'(t)$ によって配信用音楽ソフト S' が生成される。

このような手法で合成された透かし音声成分は、音楽成分とほぼ同じ帯域に広く分布するため、帯域通過フィルタなどを用いて音楽成分と透かし成分を完全に除去することが難しいと考えられる。

2.2 透かし音声除去鍵の作成法

ここでは、透かし音声を除去する鍵 K の作成法について示す。まず、音声として不可聴な状態の透かしデータ W を準備する。次に、 W の時系列標本値を $w_i(t)$ とし、その i 番目のフレームの MDCT 係数を $W_i(k)$ とする。

こうして得られた係数 $W_i(k)$ と $A_i(k)$ から

$$K_i(k) = -\beta \cdot A_i(k) + (1-\beta) \cdot W_i(k) \quad (7)$$

$$0 \leq k \leq M-1, 0 \leq \beta \leq 1$$

を生成する。ただし、 β は透かし音声データ $a(t)$ の減衰量を制御する係数である。 $K_i(k)$ を逆変換することで透かし音声除去鍵 K が得られる。ここで右辺第1項の“ $-\beta$ ”は、透かし音声の成分を相殺するための処置であることに注意する。この処理ブロックを図3に示す。

本論文では、議論を簡単にするために透かし情報 W として、知覚帯域外の低周波信号 $w_i(t)$ を用いることにした。この時系列標本値は、

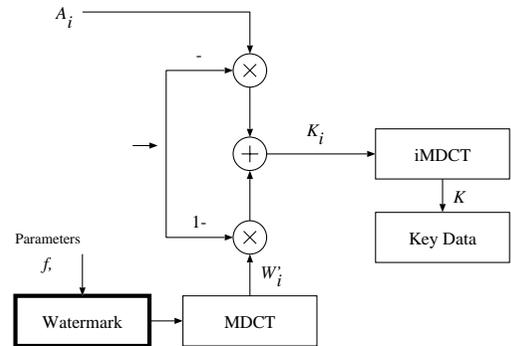


図3 鍵生成処理ブロック

Fig. 3 Key generating diagram.

$$w_i(t) = \gamma \cdot \sin\left(\frac{2\pi \cdot f \cdot t}{F}\right) \quad (8)$$

のように求められる。ただし、 f は周波数、 F はサンプリングレート、 t は標本化位置を表し、 γ は透かしの強度を制御するための係数である。人間の知覚できる周波数の下限は、一般に 20 [Hz] が限度である。これ以下の周波数音を知覚させるには 75 [dB] 以上の強いパワーが必要とされている。本研究では透かしの強度 γ を高くしても、聴者に知覚されない $f \leq 20$ とした。

2.3 鍵データの圧縮手順

現在のオンライン配信では、通信容量を削減するために高効率圧縮符号化の適用は不可欠である。音楽や音声の場合、聴感的な特性をうまく利用した不可逆圧縮が効率良くデータ量を削減できる。よって、ほとんどの圧縮手順に不可逆方式が採用されている。そこで、提案システムにデータ圧縮の機能を付加する。不可逆な圧縮方式として代表的なものに、MP3 で知られる MPEG1 Layer3 がある。ここでは、MP3 を用いて配信する試聴音楽データを圧縮するシステムについて示す。

データ圧縮を考慮した埋込みシステムのブロックを図4に示す。この図のように鍵データを MP3 で圧縮すると、透かし信号 W に大きな影響を及ぼすと考えられる。しかし、MP3 では周波数分析に用いる音声区間を最大 512 サンプルに限定しサブバンド化しているため周波数分解能は低い。この欠点を補うために MDCT を用いているが、本方式のような可聴帯域外の低周波成分は分離できない。よって、透かし信号 W は、可聴帯域の周波数成分と混在した状態で符号化されるため圧縮の影響をほとんど受けないと考えられる。一方、音楽と同じ可聴帯域に存在する透かし音声 A は、圧縮符号化によって受ける聴感的な影響は少ない

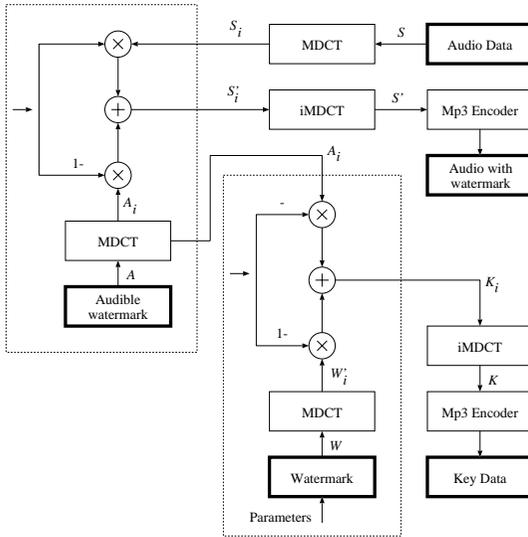


図 4 圧縮を考慮した埋込みシステム

Fig. 4 Watermark embedding diagram with compression scheme.

といえる．ただし， α を大きな値にすると MP3 の圧縮特性に影響を与え，可変透かしが制御できなくなるので α は低い値に設定することが望ましい．

3. 可変透かしの除去と秘匿透かしの埋込み

3.1 透かし音声の除去原理

試聴用音楽ソフトから透かし音声 A を取り除くことで高品質な音楽ソフト S'' を再生する原理について示す．試聴用の音楽データは，式 (4) の周波数成分 $S'_i(k)$ を持つ．これに式 (7) の音声除去鍵 $K_i(k)$ を埋め込むと

$$\begin{aligned} S''_i(k) &= \frac{S'_i(k) + K_i(k)}{\alpha} \\ &= S_i(k) + \frac{1-\alpha}{\alpha} A_i(k) - \frac{\beta}{\alpha} A_i(k) \\ &\quad + \frac{1-\beta}{\alpha} W_i(k) \\ &= S_i(k) + \frac{1-\alpha-\beta}{\alpha} A_i(k) \\ &\quad + \frac{1-\beta}{\alpha} W_i(k) \end{aligned} \tag{9}$$

$$0 \leq k \leq M-1, 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1$$

のようになる．この $S''_i(k)$ を逆 MDCT 変換して音楽データ S'' を再生する．ここで，低周波透かし成分 $W_i(k)$ は知覚できないため音声透かし $A_i(k)$ に着目すると

$$\begin{aligned} \epsilon &= \frac{1-\alpha-\beta}{\alpha} \\ 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1 \end{aligned} \tag{10}$$

で制御されることが分かる．この ϵ が音楽成分の含有率に比して低い値であるならば，透かし音声

$$A'_i(k) = \epsilon \cdot A_i(k) \tag{11}$$

は知覚できなくなる．その結果，再生音 S'' からは音楽成分のみが聴こえるため，聴感的には高品質に感じられる仕組みである．一方，再生された音楽 S'' は，新たに埋め込まれた透かし信号 W および透かし音声 A' が含まれた状態であるので， S'' が横流されたならば，秘匿透かし成分を抽出し不正行為を特定できる．本方式によれば W は K の購入後に埋め込まれるので，購入者 ID を W に混入すれば不正流通源の特定にも有用であると考えられる．購入者 ID の簡単な埋込み法として，たとえば時間軸上で区切られた音声フレームごとに透かし信号の有無を制御し，個人を識別するビット系列を埋め込む方法が考えられる．ただし，本手法をそのまま用いると帯域通過フィルタなどで容易に購入者 ID を除去される問題がある．実用化の際には， W の生成法に帯域通過フィルタなどの攻撃手法に耐えられるような電子透かし技術を導入することが望ましい．

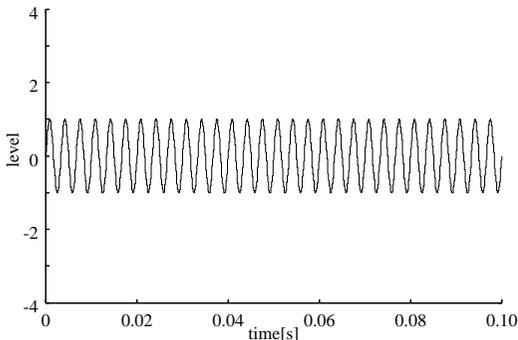
本方式では試聴用音楽データに対して不正な変換操作が施されると音声透かしの除去が困難になる．よって，試聴用音楽データに対する不正な処理を防ぐ効果も期待できる．また， $\alpha + \beta = 1$ としたならば，式 (10) と式 (11) の関係から明らかなように透かし音声 A' の影響を完全に消すこともできる．

3.2 透かし機能と検出区分

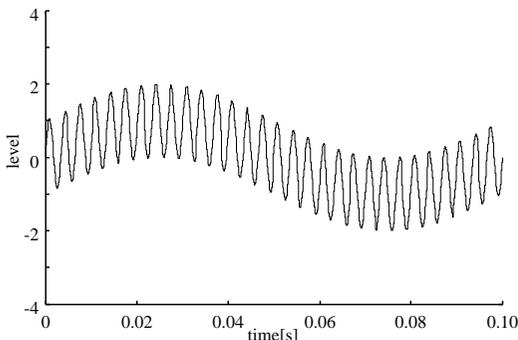
本方式の透かしによれば音楽データを表 1 のように分類できる．試聴用音楽ソフト Trial 版は自由に利用できるものとして一般に公開される．購入手続きを完了した利用者は，正規ユーザとしての ID 情報が含まれた音声透かし除去鍵を入手し Trial 版から高品質な音楽ソフト Goods 版を再生できる．このとき埋め込まれる ID 情報付きの秘匿透かしの存在によって正規利用か不正利用かを証明できる．一方，複製品である Copy 版が不正に横流されたならば，その ID 情報から流通源を特定できる．また，不正利用を目的とする信号処理によって透かし（利用者 ID）が破壊された音楽ソフト Imitation 版の使用者には，その行為に対し警告を発することができる．さらに，ID 情報の異なる複数の音声透かし除去鍵を用いた結託攻撃などにより，音声透かし成分をある程度推定し不正に除去した場合においても Imitation 版と見なせる．実用化の際には，このような不正行為から保護するための法的な対策も重要である．

表 1 透かしによるコンテンツの分類
Table 1 Classification of contents by watermarking.

	Sample use	Regular use	Illegal use	Illegal distribution except watermarks
<i>S</i>	audible	audible	audible	audible
<i>A</i>	audible	inaudible	inaudible	lost
<i>W</i>	none	inaudible	inaudible	lost
class	Trial	Goods	Copy	Imitation



(a) 300 [Hz] sound waveform.



(b) 300 [Hz] sound waveform with 10 [Hz] watermark signal.

図 5 波形からの透かし検出

Fig. 5 Check of watermarked waveform.

4. 透かしの検出法

4.1 視覚検出法

透かし信号 W として 20 [Hz] 以下の超低周波成分を用いると、透かしのスペクトル γ を強くしても知覚されない。この特徴を利用して強いスペクトルの低周波波形を埋め込み、その形状から透かし信号の存在を確認する。図 5 に 300 [Hz] の純音 S と、それに 10 [Hz] の透かし信号 W を埋め込んだ波形を示した。この純音とは単一の周波数成分のみを有する信号である。この図から W のスペクトルパワーが S に比して強ければ、低周波の存在を容易に確認できることが分かる。また、再生音からは 300 [Hz] の純音のみが知

覚されるため、図 5 (a) と図 5 (b) に聴感的な差異は感じられない。

4.2 分析検出法

音楽ソフトの利用形態によっては、波形を大きく振動させることが望ましくないことも考えられる。その場合は γ を小さな値として秘匿透かしを埋め込み、離散フーリエ変換 (DFT) による検出¹²⁾を用いればよい。

分析対象データ X の時系列標本値を $x(t)$ とし N 個の標本値を持つフレーム i のフーリエ係数 $X_i(k)$ は、

$$X_i(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \exp\left(-j \cdot \frac{2\pi \cdot k}{N} \cdot n\right) \quad (12)$$

により求まる。この計算には FFT などの高速変換アルゴリズムが利用できる。また、分析に用いる標本数 N は、サンプリングレート F と周波数成分 f によって

$$N \geq \frac{F}{f} \quad (13)$$

として定める必要がある。透かしの検出に際しては、これを考慮しなければ目的信号を抽出できない。たとえば、 $F = 44100$ [Hz] の音声から $f = 10$ [Hz] の周波数成分を抽出するには $N \geq 4410$ による分析が必要である。

4.3 差分検出法

可聴帯域中に不特定な成分として含まれる音声透かし A' は、これまでの手法では検出が困難である。そこで、音楽データ S を検出鍵として用いた差分検出法を用いる。まず、検出対象となる音楽データ S'' から得られるの MDCT 係数 $S_i''(k)$ とオリジナルの音楽データ S から得られる $S_i(k)$ を用いて、

$$\begin{aligned} \delta_i(k) &= S_i''(k) - S_i(k) \\ &= S_i(k) + \frac{1-\alpha-\beta}{\alpha} A_i(k) \\ &\quad + \frac{1-\beta}{\alpha} W_i(k) - S_i(k) \\ &= \frac{1-\alpha-\beta}{\alpha} A_i(k) + \frac{1-\beta}{\alpha} W_i(k) \quad (14) \end{aligned}$$

$$0 \leq k \leq M-1, 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1$$

表 2 実験音声

Table 2 Sound for experiments.

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

表 3 透かし音声

Table 3 Watermark sound for experiments.

Name	Samples	Sec
Message	440,832	10

を求める．この差分係数値 $\delta_i(k)$ をさらに

$$\begin{aligned} \Delta_i(k) &= \frac{\alpha}{1 - \alpha - \beta} \delta_i(k) \\ &= A_i(k) + \frac{1 - \beta}{1 - \alpha - \beta} W_i(k) \quad (15) \\ 0 \leq k \leq M - 1, 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1 \end{aligned}$$

としてスペクトルを増大し，逆変換すれば透かし音声を再生できる．この再生音から W は知覚されず聴感に与える影響は少ないが，スペクトルが音源装置の限界変域を越えないように注意しなければならない．また，再生音に混入する量子化雑音は音声透かし A に比して微少である．この手法による検出は，検出鍵としてオリジナルの音楽データが必要であることに注意する．

5. 実験結果と考察

本手法では，試聴用ソフトの再生音から透かし音声を知覚でき，その除去後に音楽ソフトの品質が極度に低下してはならない．さらに透かし音声を抑圧した後の音楽データから，密かに埋め込まれた秘匿透かしが検出できなければならない．本研究では，高品質の音楽データに透かし音声を埋め込み，その除去後の音質と透かしの検出と実験し，さらに MPEG1 Layer3 による高効率圧縮と与える影響について考察した．

5.1 実験音声および署名画像

表 2 に示した実験音声は，アナログ状態の音楽を 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．音楽ソフトは，通常ステレオ音であるが議論を簡単にするために，その片側成分のみを用いて実験した．また，表 3 に示した透かし音声は，著者による“この音声はサンプルです”という発声音を同様にデジタル化したものである．

5.2 音質の評価尺度

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

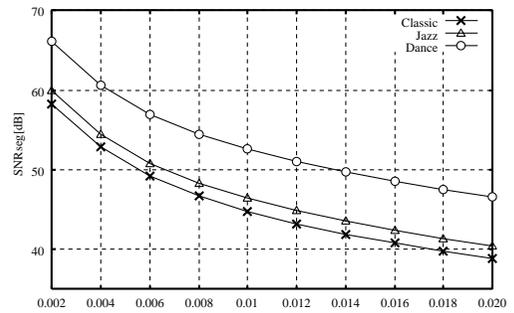


図 6 SNR_{seg} と ϵ の関係
Fig. 6 ϵ to SNR_{seg}.

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¹³⁾．

$$SNR = 10 \log_{10} \left\{ \frac{\sum_m So^2(m)}{\sum_m Er^2(m)} \right\} \quad (16)$$

本実験では，音質の評価法として SNR を改良した SNR_{seg} (Segmental SNR) を用いた¹³⁾．

$$SNR_{seg} = \frac{1}{N_f} \sum_{f=1}^{N_f} SNR_f \text{ [dB]} \quad (17)$$

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．

5.3 透かし音声の知覚限界

本方式で α は，音楽の品質を極端に劣化しない値として容易に決定できる．しかし， β は可聴な透かし音声不可聴になる値でなければならないため，その決定には注意を要する．本実験では (α, β) によって定まる ϵ と再生音質の関係について調べた．まず，(α, β) の制御により ϵ の値を変化させた再生音を複数用意した．これらのオリジナル音楽に対する SNR_{seg} を調べると図 6 が得られた．この結果から SNR_{seg} と ϵ の値には密接な関係があると分かる．また，表 2 に示した音声の場合，一般に $\epsilon < 0.016$ とすれば 40 [dB] 以上の高音質を実現できることが分かった．本実験では， $\gamma = 0$ として秘匿透かし W によって生じる波形差分値を抑制したが，低周波帯域の透かし信号を別に用意し再生しても知覚できなかった．

5.4 音声除去と不可聴透かしの検出

本手法で高音質を再現できるためには， $\epsilon < 0.016$ になる (α, β) を用いることが望ましい．実験では

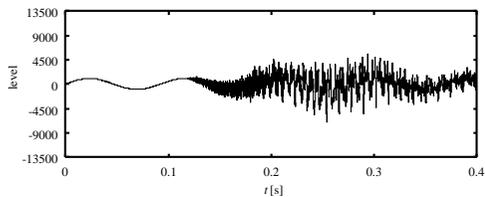


図 7 音声波形からの検出 ($\gamma = 1000$)

Fig. 7 Detection from sound waveform ($\gamma = 1000$).

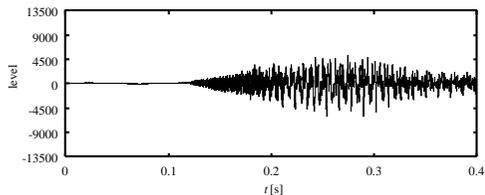


図 8 音声波形からの検出 ($\gamma = 100$)

Fig. 8 Detection from sound waveform ($\gamma = 100$).

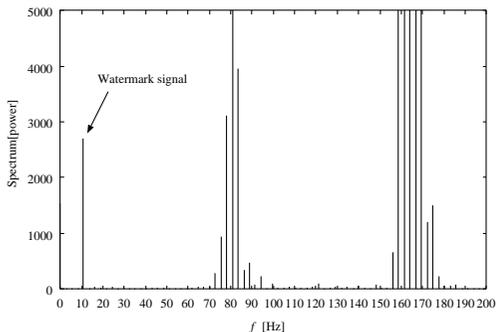


図 9 周波数分析による検出 ($N = 16384$)

Fig. 9 A result of DFT signal analysis under an insufficient number of samples ($N = 16384$).

$\alpha = 0.8$ として埋込みを施し、鍵 K の作成には $\beta = 0.19$ を用いることにした。ただし、実験音声としては表 2 の “Classic” を用い、MDCT のフレーム長は $M = 18$ とした。

まず、 $f = 10.76$ [Hz]、 $\gamma = 1000$ とした W を含む K を用いて、透かし音声を除去すると図 7 の音声波形が得られる。この波形から W の埋込みによって、強い交流成分が発生していることが分かる。さらに、その波形が W の周波数 f に一致することも確認できる。この波形の観察による判定は最も容易な検出手法の 1 つである。また、 W は可聴帯域外の信号であるため聴覚では通常知覚できないことに注意する。

次に、 $\gamma = 100$ とした場合、その音声波形は図 8 のようになる。この波形を観察しても透かし信号成分の存在を確認することは困難である。そこで、周波数分析による透かし信号の検出を試みた。図 9 は、音声データの周波数パワースペクトルのうち、低周波に

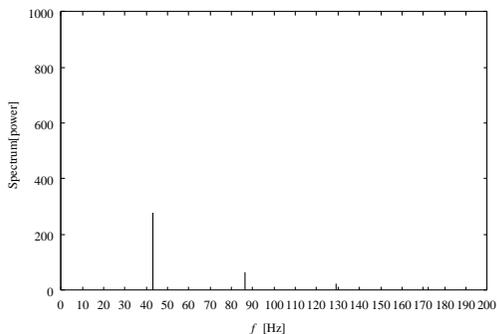


図 10 周波数分解能が不足した例 ($N = 1024$)

Fig. 10 A result of DFT signal analysis under an insufficient number of samples ($N = 1024$).

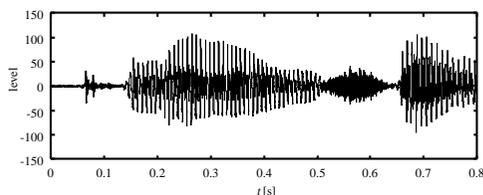


図 11 可聴帯域の透かし音声

Fig. 11 Separated watermark sound from audio data.

相当する部分を限定的に示したものである。この結果から、波形では視認できなかった W が約 10.76 [Hz] の成分として容易に確認できる。ただし、この分析には $N = 16384$ という膨大な量の標本を用いている。ここで、不十分な長さの音声区間 ($N = 1024$) を用いて分析すると図 10 が得られた。この図では W の成分が分散され、透かしの存在を特定できない。これは DFT 分析に用いた標本数とその音声区間によって、その周波数分解能が決定されることに起因する。

さらに、本方式では、音楽データと同じ可聴帯域にも透かし音声が残る工夫を施してある。そこで実験的にその抽出を試みた。図 11 に透かし音声を除去した後の音楽データからオリジナル音声を取り除いた波形を示した。この実験では波形を観察しやすくするために $\gamma = 0$ とした。この波形から透かし音声が知覚できない状態で埋め込まれていたことを確認できた。さらに、差分波形の周波数成分に $\alpha/(1 - \alpha - \beta)$ を乗じてスペクトルを増大し、再生すると埋込みに用いた透かし音声 A を聴取できた。

5.5 高エネルギー圧縮の影響

本方式が MP3 圧縮によって受ける影響について考察する。可聴帯域に埋め込まれた透かし音声は、音楽の周波数成分と同じ帯域にあるため、試聴音楽データが圧縮されても聴感的に受ける影響は少ないと考えられる。そこで、 $\alpha = 0.8$ として作成した試聴用音楽

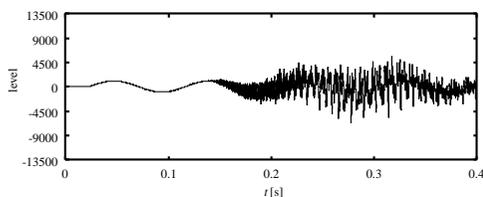


図 12 MP3 圧縮音声からの視覚的検出

Fig. 12 Visual detection from audio data compressed by MPEG1 Layer3.

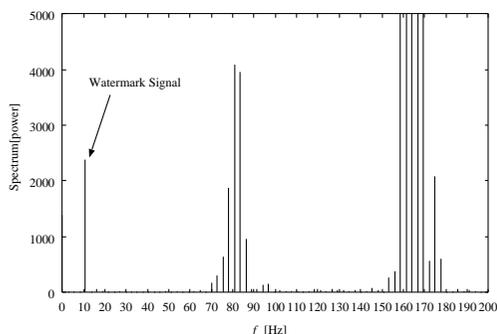


図 13 MP3 圧縮音楽からの DFT 分析検出 ($N = 16384$)

Fig. 13 Result of DFT signal analysis to audio data compressed by MPEG1 Layer3 ($N = 16384$).

データを 1/10 まで MP3 で圧縮伝送し再生したが、著者らの聴取では差異を感じられなかった。

また、鍵データの圧縮に MP3 を用いたシステムを構築し、透かし音声の除去と秘匿透かしの検出が可能かどうかを確認した。まず、 $\alpha = 0.8$ として透かしの埋込みが施された音楽データを MP3 で 1/10 まで圧縮し復元した。さらに、 $\beta = 0.19$ による鍵データを作成し、この音楽データから音声透かしを除去した。この際、 $\gamma = 1000$ とすると図 12 の再生波形が得られた。この波形から、視覚による透かしの検出は可能であることが分かる。また、 $\gamma = 100$ の低い値にした場合、波形からの検出が困難なので周波数分析して図 13 を得た。この結果から秘匿透かしのスペクトルがやや減衰するものの、その存在を十分特定できることが分かる。さらに、オリジナルの音声と同じ設定の MP3 で圧縮伸張し音楽成分を除去し再生すると、雑音が混入するものの埋め込まれた透かし音声を聴取できた。これらの結果から、本手法はオンライン配信に不可欠な圧縮符号化にも適用できると考える。

6. おわりに

本論文では、知覚できる状態の透かし音声を音楽ソフトに埋込み配信するための手法を提案した。本手法によれば、鍵データによって透かし音声のみを抑制し、

さらに不可聴な透かし信号を埋め込むことができる。また、透かし音声の除去と秘匿透かし信号の埋込みは鍵を購入した利用者によってなされる。したがって、音声除去鍵の透かし信号として密かに購入者情報を混入すれば、従来の方式では困難な不正流通者を特定する一助として有用であると考えられる。ただし、この実用化にはスペクトル拡散法などの高い秘匿性を実現できる手法⁵⁾の応用が望ましい。また、透かしと音楽信号の分離が困難な可聴帯域内にも知覚できない状態の透かしを埋め込む着意も必要である。これには、従来の反響音やマスキングを用いた手法^{4),6)}の応用が有効と考えられる。実用化に際しては、様々な電子透かしの特徴を生かして適切に応用することが重要である。

また、デジタルコンテンツの著作権を保護するためには、電子透かし技術の存在を広く認知させ、著作権者に損害を与えるような違法行為を心理的にも抑止することが重要である。本方式は、その実現に有効な一手法と考える。

参考文献

- 1) 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)。
- 2) 電子透かし技術に関する調査報告書，日本電子工業振興協会 (1999)。
- 3) 松井甲子雄，中村康弘，ナタウトサムパイブーン：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)。
- 4) 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)。
- 5) 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)。
- 6) 森本典繁：電子透かしを支えるデータ・ハイディング技術(下)，*日経エレクトロニクス*，Vol.3-10, No.684, pp.153-168，日経 BP 社 (1997)。
- 7) 富岡淳樹，中村高雄，小川 宏，高嶋洋一：マルチチャネルデ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)。
- 8) 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)。
- 9) 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).

- 10) 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).
- 11) 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).
- 12) 貴家仁志, 村松正吾: マルチメディア技術の基礎 DCT 入門, CQ 出版 (1997).
- 13) 小澤一範: デ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).

(平成 12 年 1 月 24 日受付)

(平成 13 年 2 月 1 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 年同大
学校理工学研究科情報数理専攻修了.
平成 11 年同大学校情報工学科助手.
マルチメディア符号化, 情報セキュ
リティ, 電子透かしに関する研究に従事. 電子情報通
信学会, 日本音響学会各会員.



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. 平成元年同大学校情報工学科教
授. 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「電子
透かしの基礎」(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.