

ハース効果を用いたデジタル音楽情報の 電子透かしに関する考察

山本 啓*

稲葉 宏幸*

概要

デジタルコンテンツのネットワーク配信が盛んに行われるようになり、デジタルコンテンツの著作権保護はますます重要になってきている。本報告では、心理聴覚特性の1つであるハース効果を用いることにより、聴感的に劣化の少ない電子透かし手法について提案する。計算機実験により本手法は、聴感的に劣化が少なく、MP3符号化による攻撃に対してある程度の耐性を有することを示す。

Notes on a Digital Watermarking Technique for Audio Data by Haas Effect

Kei Yamamoto*

Hiroyuki Inaba*

abstract

Recently network distribution of digital contents have been widely used, and the problem of copyright protection for digital contents becomes more important.

In this paper, we propose a new digital watermarking technique for audio data which have a high quality in auditory feeling. The new method is realized by using Haas effect which is one of the mentality auditory characters.

By computer experiment, we confirm the audio quality of the embedded audio and the tolerance to MP3 coding.

1 はじめに

近年、ブロードバンドインターネットの普及に伴い、ネットワークを介したデジタルコンテンツの配信が行われており、実際に商用化されているものもある。しかし、デジタルコンテンツは、劣化を伴わず容易に複製可能であるため、不正コピーを助長しやすい。そのため不正コピーによる著作権侵害などが問題となっている。この問題を解決するために、様々な方式が採用されてきたが、どの方式も著作権保護策としては不完全であり、完全な解決策までには至っていない。この問題を解決する一手法と

して、電子透かしが提案されている [1]。電子透かしとは、デジタルコンテンツにあらかじめ人間が知覚できないように著作権情報などの副情報を埋め込み、不正コピーを特定し、またそれらの行為を抑止する事によって著作権を保護する技術である。デジタル音楽情報を対象とした電子透かしは、画像などの視覚的なコンテンツを対象にした電子透かしに比べ、電子透かしの手法の提案数は少ない。これは、人間の感覚は視覚よりも聴覚の方が敏感であるため、実現が難しいとされているためである。

本稿では、心理聴覚特性の一つであるハース効果を用いることにより、聴感的に劣化の少ない電子透かしが実現できる事を示す。

2. では、心理聴覚特性について述べ、3. では、ハース効果を用いた電子透かしの手法を示す。5. では、本手法を用いた計算機実験の結果を示し、考察を行う。

*京都工芸繊維大学 工芸学部 電子情報工学科 情報通信システム研究室

*Department of Electronics and Information Science, Kyoto Institute of Technology

2 心理聴覚特性

2.1 ハース効果 [2]

音は壁や天井など様々なものに反射しながら順次耳に届くが、人間は一番早く到達した音の方向に音源位置を感じ、遅れて届く音を無視する特性がある。ハース効果は、一般に最初に到達した先行音と遅れて届く後続音の時間差が 1 ~ 30ms の範囲である時に生じ、時間差が 50ms を越えると分離したエコーとして後続音の存在が認められるようになる。

また、観測者から等距離にある 2 箇所の音源から同一の音が同音量で発せられた場合、観測者はそれらの音源の中央に音源位置が存在していると感じる。この時、片方の音源の位置を観測者から遠ざけると、動かしていない方の音源からのみ音が聴こえてくるようになり、遠ざけた方の音源からは音を知覚しなくなる。すなわち、動かしていない方の音源に音源位置を感じるようになる。

2.2 音の時間と強さの交換性

観測者からの距離が違う 2 箇所の音源から同一の音が同音量で発せられた場合、観測者に近い方の音源に対して音源位置を感じるが、観測者から遠い方の音源の音量を上げていくと、次第に音源位置が観測者に近い方の音源から遠い方の音源へと移って行く。

これにより、観測者からの距離が違う 2 箇所の音源から同一の音が発せられた場合でも、音量が異なっていれば、音像の位置は音源を観測者から等距離にした場合の中心位置にくる。

これを 2.1. で述べたハース効果と併せて考えると、音の時間と強さには交換性があることが分かる。

これは、刺激が加わってから神経が発火するまでの遅れ時間である神経の潜時と関係があると考えられている。刺激が加わって神経が応答するまでには多少の時間遅れがあるが、潜時は全ての刺激に対して一定時間ではなく、刺激が強ければ短く、刺激が弱ければ長くなるのが普通である。このため、強度の弱い音と強度の強い音をある時間だけ遅らせて聞いたとしても、神経が強度の強い音に対して先に応答すれば、先に強度の強い音を聞いたように感じる。

3 提案手法

2. で述べた心理聴覚特性を利用した電子透かしを提案する。ハース効果を利用すれば、知覚されにくい音を付加、または、削除する事によって透かし情報を埋め込み、それにより主観的な音質の劣化が少ない電子透かしを実現できるものと思われる。

3.1 電子透かしの埋め込み方法

デジタル音楽情報は、音楽の性質上そのほとんどがステレオ音楽であると考えられるので、ステレオ音楽に適した電子透かしの手法を考える。

2. で述べた通り、左右チャンネル信号間で周波数成分の振幅に大きな差がある部分では、周波数成分の振幅が小さい音は知覚されにくいので、この部分に対して電子透かしを埋め込む手法を提案する。

この手法における電子透かしの埋め込みのブロック図を図 1 に示す。図 1 のハイブリッドフィルタバンクは Mpeg Audio Layer - 3(MP3) で用いられているものと同じである [5] [6] [7]。電子透かしの埋め込みは、MDCT ブロックの後に埋め込みブロックを設け、MDCT 係数を変更することにより行う。

また、MP3 に対する耐性評価を行いやすくするため、埋め込み処理は MP3 で用いられているフレーム長毎に行う事とする。

以下に電子透かしの埋め込みアルゴリズムを示す。

1. 埋め込みを行うフレーム数を f 、左右チャンネルの MDCT 係数を各々 $l(n)$, $r(n)$ 、左右チャンネル間の比 $R_{LR}(n)$ ($n = 1 \sim 1152 \times f$) を次式により求める。

$$R_{LR}(n) = (\text{Sign}(l(n) \cdot r(n))) \cdot \frac{\min(|l(n)|, |r(n)|)}{\max(|l(n)|, |r(n)|)} \quad (1)$$

$$(n = 1, 2, \dots, 1152 \times f)$$

上式で、 $\text{Sign}(a)$ 関数は、 a が正の時は 1、 a が負の時は -1 を返す関数とする。

2. $R_{LR}(n)$ と閾値 $\epsilon (> 0)$ とを比較して以下の条件に従って処理を行う。

- (a) $R_{LR}(n) \geq \epsilon$ または $R_{LR}(n) < 0$ の場合埋め込みを行わない。

(b) $0 \leq R_{LR}(n) < \epsilon$ の場合

埋め込む透かしビットを $b \in \{0, 1\}$, 埋め込み強度を w とし,

i. $|l(n)| < |r(n)|$ の場合

$$l(n) = w \cdot b \cdot l(n) \quad (2)$$

ii. $|l(n)| > |r(n)|$ の場合

$$r(n) = w \cdot b \cdot r(n) \quad (3)$$

により埋め込みを行う。

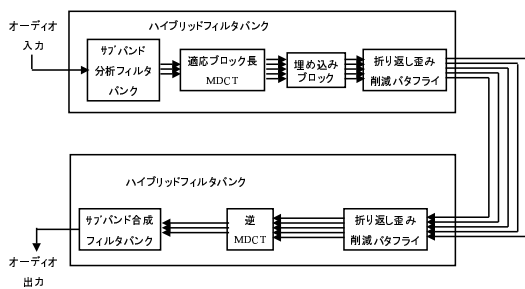


図 1: 提案手法の埋め込みブロック図

(a) $\mu = 0, \sigma^2 = 0$ の場合

無音部分であり透かしビットの埋め込みが行われていないと判断して、無視する。

(b) $\mu \leq \epsilon'$ の場合

抽出した透かしビットを 0 とする。

(c) $\mu > \epsilon'$ の場合

抽出した透かしビットを 1 とする。

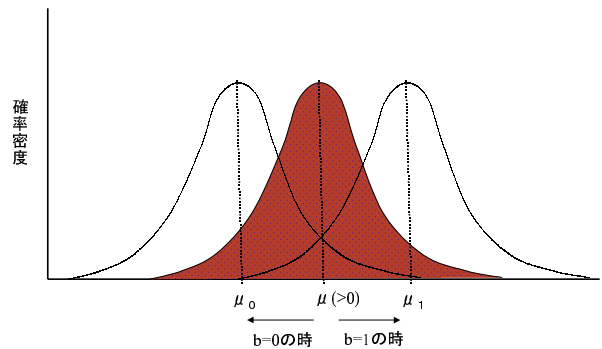


図 2: $R_{LR}(n)$ の分布の変化

3.2 電子透かしの抽出方法

以下に透かしビットの抽出アルゴリズムを示す。

1. 埋め込みと同様の手順で $R_{LR}(n)$ を求める。(式 (1) 参照)
2. $R_{LR}(n)$ の平均 μ と分散 σ^2 を以下のようにして求める。

$$\mu = \frac{1}{N} \cdot \sum_{i=1}^n R_{LR}(n) \quad (4)$$

$$\sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^n (R_{LR}(n) - \mu)^2 \quad (5)$$

なお, $N = 1152 \times f$ である。

3. 透かしビットを埋め込んだ $R_{LR}(n)$ の分布は, 3.1 の埋め込みによって, 図 2 のように移動しているはずである。そこで, 求めた平均 μ と分散 σ^2 と閾値 ϵ' を用いて以下のように透かしビットの抽出を行う。

4 音質評価法 [8]

従来は, 音質の客観的な測定尺度として SNR が広く使用されてきた。しかし, MP3 などの聴覚のマスク特性を利用した低ビットレートで高音質な音響信号符号化器の音質を測定するためには, SNR は適さない。

そこで, 本稿では SNR に代わる客観的な音質評価法として, ITU-R が作成した PEAQ(Perceived Evaluation of Audio Quality) 方式を用いて評価を行う。この客観評価法は, 原音と高能率符号化音などの処理音の 2 信号を用いて, 処理音の音質を測定する。この評価法の特徴は, 主観評価法にそって行った主観評価テストに対応した評点の測定が可能であり, 0 から -4 の範囲で音質を客観的に測定する。ここで, 0 は, 処理音の音質は原音と同じで, 全く劣化していないレベルを表し, -4 は, 処理音の音質の劣化が非常に気になるレベルを表している。

5 計算機実験と考察

計算機実験は、サンプリング周波数 44.1kHz、量子化ビット数 16 ビット、22 秒、857 フレームの一般的なデジタル音楽情報を対象として行った。実験パラメータは、 $f = 5$ 、埋め込みビット数 20 ビットとし、 w 、 ϵ 、 ϵ' の各々の値を変化させながら、埋め込みを行った。

3.1 の手法を用いて 20 ビットのデータを繰り返し埋め込むことにより、透かし入り音楽データを作成した。実験に用いたサンプルの場合には 20 ビットの各埋め込みビットを 8 ~ 9 回埋め込むことが可能であった。透かし情報の復号は、3.2 の手法を用いて復号した結果を多数決復号することにより行なっている。

5.1 音質評価

埋め込みを行った後の音質評価を以下の図 3 に示す。ここで、図 1 のブロック図において埋め込みブロックを使用しなかった場合の音質評価値は、-3.03 であった。この値と図 3 を比較すると、埋め込みによる劣化は少ないと考えられる。筆者が試聴する限りでは、主観的な音質の劣化は感じられない程度であった。音質評価の値があまり良くなかったのは、処理プログラムの埋め込みブロック以外の処理ブロックの影響によるものと考えられる。

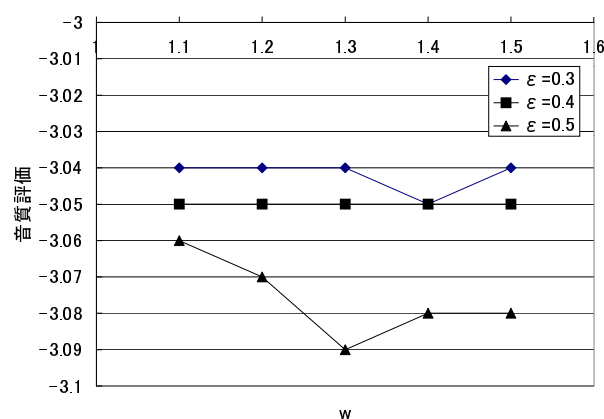


図 3: 埋め込み後の音質評価

5.2 MP3 符号化攻撃

MP3 符号化攻撃後の透かしの抽出率を図 4 に示す。この図より、提案した電子透かしの手法は、MP3 に対してある程度の耐性を有することがわかる。

図 3、図 4 より、音質の劣化と電子透かしの耐性は、トレードオフの関係になっているが、図 3 より、 ϵ や w が音質評価値に与える影響はわずかであることがわかる。

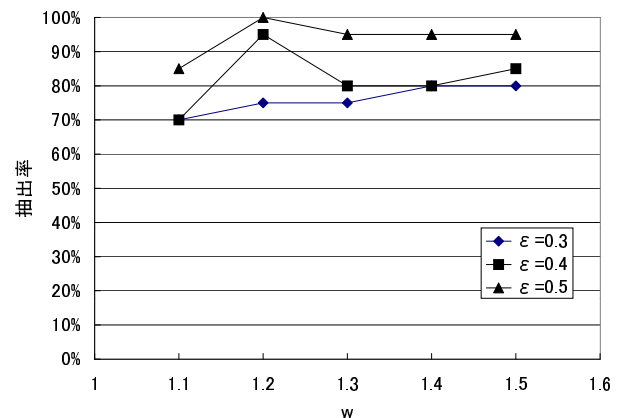


図 4: MP3 符号化攻撃後の抽出率

6 むすび

本稿では、ハース効果を用いてデジタル音楽情報に電子透かしを埋め込む手法を提案した。本手法は、主観的な音質の劣化がほとんどなく、MP3 符号化攻撃に対してもある程度の耐性を有することを示した。今後の課題としては、音質劣化を抑えるための方式の改良と、サンプリング周波数の変換や再量子化など MP3 符号化以外の攻撃に対する検討、また、埋め込み時の最適なパラメータの決定などが挙げられる。

なお、本研究の一部は日本学術振興会科学研究費補助金基盤研究 B(1)124501571 による。

参考文献

- [1] 松井甲子雄:電子透かしの基礎, 森北出版 (1998)

- [2] 日本音響学会・境久雄, 中山剛:聴覚と音響心理, コロナ社 (1978)
- [3] 川人光男, 行場次郎, 藤田一郎, 乾敏郎, 力丸裕:岩波講座 認知科学 3 視覚と聴覚, 岩波書店 (1994)
- [4] 樋渡涓二, 安田稔, 大串健吾, 斎藤秀昭:視聴覚情報概論, 昭晃堂 (1987)
- [5] テレビジョン学会:総合マルチメディア選書 MPEG, オーム社 (1996)
- [6] 中村新一:MP3 の符号化技術の基礎, Interface 8月号, CQ 出版社 (2000)
- [7] 小杉篤史:MP3 と等価的なエンコーダとデコーダの紹介, <http://www5c.biglobe.ne.jp/atsu/>
- [8] 渡辺馨:ITU-R 勧告に準拠した客観音質評価法, 画像電子学会第7回 VMA 研究会 (2001)