

心理音響モデルを用いたオーディオ電子透かし法の改良

Improvement of Digital Watermarks for Audio Signal Based on Psychoacoustic Masking Model

植村 優†
Masaru Uemura

田中 章†
Akira Tanaka

河口 万由香†
Mayuka F. Kawaguchi

宮腰 政明†
Masaaki Miyakoshi

1 序論

電子透かしは画像や音楽といったデジタルコンテンツにデータを埋め込む技術であり、著作権保護やコンテンツの不正利用防止に有効な手段だと考えられている。電子透かしには、透かしの存在が知覚されないこと、攻撃に対して耐性を持つこと、十分な量のデータを埋め込み可能であること等が要求され、一般にこれらの性能はトレードオフの関係にある。特にオーディオ信号に関しては、人間の耳はわずかな音の違いにも敏感に反応するため聴覚特性を十分考慮した技術が求められる。

オーディオ信号への電子透かし法として、スペクトル拡散を用いる方法 [1]、特定の周波数領域を加工する方法 [2, 3] 等が考案されており、それぞれ長所と短所がある。心理音響モデルを用いた Boney らの手法 [4] は音質、耐性の点で有望であるが、検出に原信号と適切な閾値設定が必要であるという問題がある。本論文では Boney らの手法に着目し、埋め込みに工夫を施すことによりこれらの問題の解決を図る。また数値実験によってその有効性を検証する。

2 既存手法

Boney らの手法 [4] では、心理音響モデルを用いてマスキングレベルを計算し、そのレベル以下になるような雑音を透かし信号として埋め込む。以下でその概要について述べる。

2.1 心理音響モデルを用いたマスキングレベルの計算

人間の耳には、大きな音の周囲の小さな音は聞こえない(同時マスキング)、あるレベル以下の音は静かな環境下でも聞き取れない(最小可聴限界)等の聴覚上の特性がある。これらの特性をモデル化したものが心理音響モデルであり、そこから計算される、人間が聞くことが出来る音量の最小値をマスキングレベルという。心理音響モデルの考え方は、mp3に代表される音楽信号の不可逆圧縮に積極的に用いられている。

2.2 情報の埋め込み

原音楽信号を N サンプル毎のセグメントに分け、それぞれについて以下の手順で情報の埋め込みを行う。

1. N サンプルのデータから心理音響モデルを用いてマスキングレベル M を計算する。
2. 長さ N の M 系列 $m(t)$ (ランダムなビット列) を発生させ、 M を周波数応答とするフィルタを通すことによりマスキングレベル以下のランダム系列 $w(t)$ を作成する。
3. $w(t)$ を透かし信号として時間軸上で原信号 $s(t)$ に加算して、埋め込み済みの信号 $y(t)$ を得る。

上記の操作を最後のセグメントに到達するまで行う。 $w(t)$ は M を任意のレベルで下回るように重み付けする。このとき、 $w(t)$ が M に近いほど埋め込み強度が高いという。又、中山ら

の方法 [5] に従い、ダウンサンプリングや信号圧縮に対する耐性を強くするため、低域通過フィルタを用いて $w(t)$ の帯域を 12kHz 以下に制限する。

2.3 情報の検出

埋め込み済み信号 $y(t)$ に対し、何らかの攻撃や量子化誤差の影響を受けた信号を $y'(t)$ とする。検出の際は $y'(t)$ から原信号を差し引き、その差分と、原信号から計算される $w(t)$ との相関係数を計算する。相関係数を $c = c\{y'(t) - s(t), w(t)\}$ と書くと、 c がある閾値 r を超えていれば、透かし信号 $w(t)$ が埋め込まれていたと判断する。セグメントサイズ N と、埋め込みに用いた M 系列が検出の鍵となる。

2.4 既存手法の問題点

Boney の手法では、検出時に閾値 r と原信号 $s(t)$ が必要である。 r として適切な値を選ばないと、透かしの検出率が低下する。既存の研究では r をその都度試行錯誤的に定めており [3, 4, 5]、実用上問題となり得る。又、原信号が必要であるため、例えば不正コピー防止等、原信号無しで検出すべき用途に用いることが出来ない。

3 提案手法

3.1 閾値を不要にする方法(提案手法 1)

従来研究 [4, 5] では、音楽信号に具体的にどのようにしてデータ(例えば文字列)を埋め込むかは議論されてこなかった。ここで、反応検査方式 [6] の考え方を用い、検出を二択の問題に帰着させることを考える。すなわち、 N サンプルのセグメント毎に 0 又は 1 の 1 ビットの情報を埋め込むものとし、0, 1 には異なった 2 つの系列をあらかじめ関連づけておく。セグメント i 番目の M_i 以下の透かし信号の計算を $w_i(t) = M_i\{m_i(t)\}$ と書くと、埋め込みと検出は以下の式で表される。

[埋め込み]

$$w_i^1(t) = M_i\{m_i^1(t)\}, w_i^0(t) = M_i\{m_i^0(t)\}$$

$$y_i(t) = \begin{cases} s_i(t) + w_i^1(t) & (\text{if } b_i = 1) \\ s_i(t) + w_i^0(t) & (\text{if } b_i = 0) \end{cases} \quad (1)$$

[検出]

$$c_i^1 = c\{y_i^1(t) - s_i(t), w_i^1(t)\}, c_i^0 = c\{y_i^0(t) - s_i(t), w_i^0(t)\}$$

$$b_i = \begin{cases} 1 & (\text{if } c_i^1 \geq c_i^0) \\ 0 & (\text{if } c_i^1 < c_i^0) \end{cases} \quad (2)$$

このように決めておけば相関係数の大小関係から埋め込みビットを判別出来るので、閾値設定の必要性が回避される。 w^1 と w^0 の相関が低いほど区別がつけやすいので、 m^1 と m^0 も互いに相関が低くなるように選ぶと良いと考えられる。今回は 1 と -1 が逆転した 2 系列(相関係数が -1)を用いる。

3.2 原信号を不要にする方法(提案手法 2)

ステレオ信号の左右にある程度の類似性があることに着目し、一方に透かしを埋め込み、もう一方を原信号とみなして検

†北海道大学大学院情報科学研究科 CS 専攻

出を行う。K個のセグメントを1ブロックとし、左右のチャンネルのうち片方を埋め込み用信号 $s^b(t)$ 、他方を検出用信号 $s^d(t)$ として、各ブロック毎に以下の手順で埋め込みを行う。

1. $s^b(t)$ の K 個のセグメントに対し 3.1 節の方法で 0,1 を試験的に埋め込み、 $s^d(t)$ を検出用信号として検出を行い、正しく検出可能なセグメントかどうかを検査する。
2. 各セグメントの左右の信号の類似度を計算する。
3. 類似度の高い方からいくつまでのセグメントで正確な埋め込み、検出が可能かを判定する。
4. 得られた埋め込み個数 p をそのブロックの検出鍵として出力し、埋め込み可能なセグメントに時系列順にデータを埋め込む。

検出の際は、ブロック毎に、左右の類似度が高いほうから p 個のセグメントに埋め込みが行われたと判断し、3.1 節の手順で 0,1 を区別する。類似度には“左右の信号の差分のノルムの小ささ”を用いる。又、埋め込み量が減るため、3.1 節で示した周波数帯域の制限は行わない。

4 評価実験

提案手法の有効性と影響を検証する。原信号として 10 秒の音楽信号を用い、 $N = 512$ 、埋め込むデータは 0 と 1 が等確率で現れるものとした。

4.1 提案手法 1

透かしが埋め込まれていると判断されたセグメントを 1、されなかったセグメントを 0 として従来通り閾値を用いて検出を行う方式と、提案方式で埋め込みを行った信号に mp3 圧縮を施した後検出を行い、得られた相関係数を埋め込んだデータで区別したものを図 1、図 2 に示す。この結果より、提案手法の相関値がよく分離されていることがわかる。

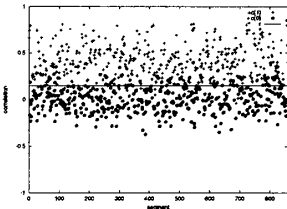


図 1 相関値と適切な閾値

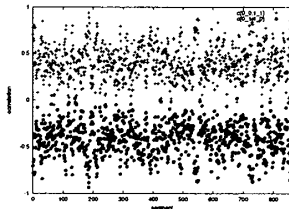


図 2 提案手法の相関値

次に、閾値を用いた方法、提案手法 1 に対し各種の攻撃を行った場合の検出率を表 1 に示す。ただし、閾値は検出率が高くなるように試行錯誤的に定めた。

表 1 攻撃を加えた後の検出率

	mp3 圧縮		LPF
	64kbps	128kbps	
既存手法	87.7%	99.9%	89.9%
提案手法	99.7%	100.0%	97.7%

4.2 提案手法 2

左右が比較的類似したステレオ信号二種類に、マスキングレベルより -40dB の強度で埋め込みを行い、埋め込み量を調べた。その後、更に各種攻撃を加えて検出率を調べた。結果を表 2 に示す。

表 2 二種類の音楽への埋め込み率と攻撃を加えた後の検出率

	埋めこみ率	mp3 圧縮		LPF
		128kbps	256kbps	
信号 1	35.3%	50.7%	90.4%	89.7%
信号 2	93.6%	55.8%	99.9%	98.9%

4.3 考察

閾値が不要な形で検出することにより、高い検出率を得ることが出来た。埋め込み強度をマスキングレベルより 20dB 以下とした場合は聴覚的にほとんど劣化が感じられなかったが、 10dB 以下にするとはっきりわかるノイズが感じられた。 $N = 512$ としたため埋め込み量は約 86bit/sec だが、用途に応じて、埋め込み量を犠牲にして N を増やすことにより、更に検出精度を上げることが出来る。

原信号を無くした場合は、信号により埋め込み量に大きな差があり、耐性も十分とは言えない。考えられる原因として、ある程度の埋め込み量を達成するために周波数帯域の制限を行っていないこと、音質を保つために埋め込み強度を低く設定していることが考えられる。又、 K が小さいほど透かしを埋め込みやすくなるが、それに伴って検出鍵の量も増えることになる。

5 結論

Boney の電子透かし法から、検出時の閾値と原信号を不要にして任意のデータを埋め込む方法を提案した。今回は二つの系列を用いたため、セグメントあたりの埋め込み量は 1 ビットだが、相関が低い 2^n 個の系列を用いれば、セグメントあたり n ビットの情報を埋め込むことも可能だと考えられる。その場合の耐性、検出精度については更に実験を行う必要がある。

原信号を不要とする方法については、埋め込み量を制限し、対象をステレオ信号に限定することにより実現出来たが、埋め込み量、耐性を安定させるための最適な類似度や、どのような音楽に多くのデータが埋め込み可能かを検証する必要がある。

参考文献

- [1] 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, Sep.1998.
- [2] 池田幹男, 豊嶋亮蔵, 武田一哉, 板倉文忠, “バンドエリミネーションを用いたオーディオ信号へのデータハイディング,” 信学論 D-II, vol.J83-D-II, no.11, pp.2350-2359, Nov.2000.
- [3] 村松巖, 荒川薫, “オクターブ類似性に基づくオーディオ信号への電子透かし,” 信学論 A, vol.J87-A, no.6, pp.787-796, June.2004.
- [4] L.Boney, A.H.Tewfik, and K.N.Hamdy, “Digital watermarks for audio signals,” European Signal Proc. Conf., Trieste, Italy, Sep.1996.
- [5] 中山彰, 陸金林, 中村哲, 鹿野清宏, “心理音響モデルに基づいたオーディオ信号の電子透かし,” 信学論 D-II, vol.J83-D-II, no.11, pp.2255-2263, Nov.2000.
- [6] 岩切宗利, 松井甲子雄, “反応検査方式を用いたデジタル音楽への電子透かし,” 情報処理学会論文誌 vol.43, no.8, pp.2519-2528, Aug.2002.