

MDCT 係数操作によるオーディオ電子透かしの音質劣化抑制 Noise Suppression due to Alteration of MDCT Coefficients for Audio Watermarking

高木 幸一†
Koichi Takagi†

酒澤 茂之†
Shigeyuki Sakazawa†

滝嶋 康弘†
Yasuhiro Takishima†

1. まえがき

オーディオ信号に対して電子透かしを挿入する方式として周波数領域における信号操作を行う方式が多数検討されている[1]。本稿では特に、MP3[2]やAAC[3]に代表される圧縮済みのオーディオデータに対して、そのMDCT係数を操作する方式を取り上げる。

オーディオ符号化におけるMDCTの特徴として、時間領域における1つのframe(単位時間ごとのサンプルの集合)が2つの隣接するMDCT係数列のframeのオーバラップの結果求められるため、MDCT係数操作の影響は時間領域で2つの隣接するframeに及ぶことになる。

そこで本稿では、透かし挿入のために、あるMDCT係数を操作したときに、それと隣接するframeに存在するMDCT係数を適応的に操作することにより、MDCT係数上で伝搬するノイズ成分を抑制する方式を提案する。

2. 提案手法とその原理

2.1. 電子透かし表現のためのモデル

提案方式が対象とする電子透かし方式はMDCT係数の操作を用いるものであり、かつ、その操作対象となるMDCT係数は比較的まばらに選択されたオーディオframeから複数選択されたものである。

操作を加えるframe(インデックスを*i*で表す)において、MDCT係数のペアの相対的な大小関係により透かし情報を表現する。N組のMDCT係数のペアの操作によりNbitの情報挿入が可能である。ここで、その中の係数ペアのひとつとして、*x*番目、*y*番目のMDCT係数のペアが選択されるとして、これらをそれぞれ($C[x,i]$, $C[y,i]$)とする。ここで、この係数ペアに挿入される情報を $E[i,x,y]$ とした場合に、により情報の挿入・検出を行うものとする。

挿入側の処理として、*i*番目のframeのある係数のペア

$$E[i,x,y] = \begin{cases} 0 & (|C[x,i]| \geq |C[y,i]|) \\ 1 & (|C[x,i]| < |C[y,i]|) \end{cases} \quad (1)$$

($C[x,i]$, $C[y,i]$)によって情報'0'を表現する際に、 $|C[x,i]| \geq |C[y,i]|$ であればこれらの係数のペアに対する操作は行わない。逆に $|C[x,i]| < |C[y,i]|$ であった場合は、不等号を逆転させるために適切な値を2.3節に述べるように $C[x,i]$ から減算する。

(一方で検出側は、埋め込み側における処理同様、*i*番目のframeのMDCT係数列の係数のペア($C[x,i]$, $C[y,i]$)の大小に着目し、式(1)に基づき、電子透かしデータ $E[i,x,y]$ の抽出を行う。)

† (株) KDDI 研究所
KDDI R&D Labs. Inc.

2.2. 電子透かし挿入による影響

前節のモデルに基づきMDCT係数に直接電子透かしを挿入する際には以下の点に注意する必要がある。

- MDCT係数の計算と時間領域でのノイズ伝搬

時間領域における1つのframeは2つの隣接するMDCT係数列のframeのオーバラップの結果求められるため、MDCT係数操作の影響は時間領域で隣接するframeに及ぶ。

- テンポラルマスキングの影響

人間の聴覚特性の一つとして時間方向のマスキング効果(テンポラルマスキング)が知られている。これは、特に時間軸上において、ある周波数の強い音が存在した場合、その前後の時間における同一周波数の音はかき消されるという性質である。特に直後の音に対してはその影響が強く及ぶことも知られている。

以上から、透かし挿入に伴うノイズ抑制の検討にあたっては、MDCT係数の原音からの乖離を相殺する際に、テンポラルマスキングにより検知限以下となる音があることに留意する必要がある。

2.3. MDCT係数操作ならびにノイズ補償方式

本節では、前節の留意点を踏まえ、提案手法であるMDCT係数操作によるノイズの影響を補償する方式について述べる。2.1節のモデルを想定した場合、その操作による影響を補償するために、下線部に相当するMDCT係数の操作方法、並びにその操作に対する補償を行う方法として以下を適用する。

【係数操作方法】前節の不等号を逆転させるための処理として、 $C[x,i]$, $C[y,i]$ のペアで絶対値が大きいほうからその変数の符号を含めた定数を減じて不等号を逆転させる。例えば、 $E[i,x,y]=0$ かつ $|C[x,i]| < |C[y,i]|$ の場合は $C[y,i]$ の符号を $\text{sign}(C[y,i])$ とした場合に"sign(C[y,i]) × (正定数)"を減ずる。なお、この正定数は2.1節下線部の目的を達成するものであればどのような値でもよい。逆に $E[i,x,y]=1$ かつ $|C[x,i]| \geq |C[y,i]|$ の場合は $C[x,i]$ から同様に定数を減ずる。

【補償処理方法】定数を減じたMDCT係数と同一のインデックスを持つ直前のframeの係数、すなわち、 $C[x,i-1]$ あるいは $C[y,i-1]$ に対し、上記で減じた値、すなわち"sign(C[y,i]) × (正定数)"を加える。

以上の方法により、MDCT係数の操作結果の伝搬による隣接frameでのMDCT係数値乖離を抑制している。さらに、直前frameの係数の絶対値を大きくすることで、マスキング効果により透かしが埋め込まれる次frameに対しその影響を与えることができ、ノイズの影響を最大限抑制することが可能な方式となる。

3. 評価実験

前節の提案手法の有効性を確認するために主観・客観の双方の評価実験を行う。なお本稿では、複数の音源(Pops, Jazz, A cappella)を MPEG-4 AAC, 44.1kHz, 96kbps で符号化したデータに対し

- (a) 提案手法 (1frame おきに電子透かしを挿入し、さらに 2.3 節に述べた補償方法に基づき直前の frame の係数を操作する方法)

の評価を行う。また、比較対象として

- (b) 全frameに電子透かし信号を挿入する場合¹
 (c) 1frame おきに電子透かし信号を挿入する場合
 (d) 1frame おきに電子透かし信号を挿入し、さらに提案手法同様に隣接 frame の操作を行うが、その操作を直前ではなく直後の frame に適用する場合。

に対する評価も行うこととする。なお、本実験において、電子透かしは各 frame 事前に決められた低域に 2つ、中域に 1つの計 3つの係数に対しランダムな値(0/1)を埋め込んだ。また、2.3 節の(正定数)としてはそれぞれの係数について 2.1 節の下線部の目的を達成するための最小の正定数を用いた。

3.1 主観評価実験

ここでは、勧告[4]を用いた主観評価実験を行う。これは “double-blind triple-stimulus with hidden reference” と呼ばれる方法で、各評価対象音に対する電子透かしの埋め込み方法を被験者（被験者は 25 人使用）には伏せた上で、基準音からの劣化度（乖離量）を測定するものである。なお、基準音は電子透かしが埋め込まれていない符号化音である。さらに、基準音のコピーも評価対象音の一つとしたため、Difference grade, すなわち “(各電子透かし埋め込み音に対する評価値) – (基準音のコピーに対する評価値)” を求めることとする。さらにその結果に対し検定を行い、片側 5% のデータを棄却し平均を求めた結果を表 1 に示す。

表 1：主観評価実験結果(grade)

	Pops	Jazz	A cappella
(a)	-0.856	-0.303	-0.068
(b)	-0.976	-0.336	-0.076
(c)	-0.872	-0.295	-0.160
(d)	-1.004	-0.305	-0.108

同表より、(a)提案方式は他方式と比較しほぼ同等、あるいは優れた結果を示している。特に今回は低域に電子透かしを埋め込んだため、Jazz を除いた人間の声が挿入された 2 音源に関し、他方式に対する提案手法の結果に明確な優位性が現れていることがわかる。特に A cappella について、単に 1frame おきに電子透かしを挿入する場合(方式(c))に比較し提案手法を適用する効果が大きいことがわかる。一方で、提案手法を除いた 3 方式を比較すると音源により優位な方式が異なる。特に(d)が全方式の中で一番悪い結果を示しうることに注意が必要である。

¹ 他条件と比較すると 2 倍の電子透かし信号が挿入されることになるが、参考のために比較対象の 1 つとして提示している。

3.2 客観評価実験

心理聴覚モデルを用いた評価尺度である勧告[5]を利用し、客観評価を行った。本勧告は、各評価対象音と基準音を比較し、乖離量を前節の主観評価指標でも用いられている grade に相当する ODG(Objective Difference Grade) 値で評価するものであり、大きな値(0 に近い値)ほど基準音との差が僅少であることを意味する。なお、基準音としては前節同様に符号化されたデータを用いることとする。以上の結果を表 2 に示す。

表 2：客観評価実験結果(ODG)

	Pops	Jazz	A cappella
(a)	-2.612	-3.628	-3.446
(b)	-3.888	-3.851	-3.780
(c)	-3.832	-3.595	-3.444
(d)	-3.766	-3.892	-3.860

以上の結果から、3.1 節同様、(a)提案手法が他方式とほぼ同等、あるいはそれ以上の結果を示していることがわかる。特に Pops において(a)提案手法の他 3 方式に対する優位性が顕著に現れている。これは、同評価音が人間の声と楽曲の両方を含んだデータであることから、透かしを埋め込んだ位置に起因する結果であると考えられる。一方で、その他 3 方式をそれぞれ比較しても音源により優位な方式が異なっている。

3.3 考察

前節の各評価実験の結果から、

- 音源の周波数の広がりと挿入する電子透かしの位置により提案方式の有効性は明確に現れるため、本方式の最適な適用を行いうためには電子透かしの挿入位置は音源に依存して適応的に決定しなくてはならない。
- (a)と(d)の比較から、補償処理は透かし挿入フレームの前 frame に対し行う方がよいことがわかった。これはテンポラルマスキングによる効果である。

と結論づけられる。

4. おわりに

本稿では、オーディオ信号への透かし挿入において MDCT 係数を操作したときに、それに伴う音質の劣化を防ぐための手法として、特にテンポラルマスキングを考慮した上で隣接 frame の MDCT 係数を操作する方法を提案した。実験の結果、主観・客観両評価の上で本方式の有効性を確認することができた。

今後は、MDCT 係数を利用した複数の電子透かし手法に本方式を適用し、音質、圧縮オーディオファイルに対する挿入・検出の高速化の観点に加え、電子透かし強度の観点からも有効な手法に関する検討を行う。

参考文献

- [1] http://www.jpo.go.jp/shiryousonota/hyoyoun_gijutsu/denshi_sukashi/2_a_1_2.htm
- [2] 例えば ISO/IEC 13818-3:1998, “Generic coding of moving pictures and associated audio information –Part 3: Audio”.
- [3] 例えば ISO/IEC 14496-3:2001, “Coding of audio-visual objects – Part 3 Audio”.
- [4] ITU-R BS. 1116-1, “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems”, Oct.1997.
- [5] ITU-R BS. 1387-1, “Method for objective measurements of perceived audio quality”, Nov. 2001.