

原曲をピッチ変換した音を用いた音楽電子透かし法

An Audio Watermarking Method Using Pitch-Transformed Sound of Original Music

村田 晴美†
Harumi Murata

荻原 昭夫‡
Akio Ogihara

1. まえがき

デジタルコンテンツに対して他のデジタル情報を埋め込む技術として電子透かしがある。音楽を対象とした電子透かし法では、ステゴ信号の音質が劣化しないことが求められている。既存の手法では、ホスト信号の成分を操作して埋め込んでおり、埋め込みに起因するノイズが知覚された場合、不快な音となる。

本稿では、ステゴ信号が音楽として成立する音質であれば、ホスト信号と異なる音が知覚された場合であっても問題はないという考えに基づいて透かしを埋め込む。そのために、ホスト信号をピッチ変換した音を透かし信号とし、透かし信号の成分を操作する。これにより、ホスト信号の成分を操作することなく透かしを埋め込むので、音質の向上が期待できる。さらに、文献[1]の手法を参考にし、位相の符号を考慮して透かしを埋め込む。

2. 原曲をピッチ変換した音を用いた音楽電子透かし法

ホスト信号と似た音色をもつ音としてサンプリング音を用いた手法を提案している[2]。しかし、ホスト信号と同じ楽器音を使用した場合であっても音色は必ずしも一致しないので、知覚された場合に違和感がある。そこで、ホスト信号そのものを加工した音を透かし信号とし、ホスト信号との音色の違いを少なくすることで音質の向上を図る。本稿では、SoundEngineFree ver5.02 [3]を用いてホスト信号をピッチ変換した音を透かし信号とした。

2.1 透かしの埋め込み

ホスト信号 $x(n)$ を連続した長さ N のフレームに分割する。各フレームに離散フーリエ変換(DFT)を施して式(1)の $X_i(k)$ ($k=0, 1, \dots, N-1$) を得る。

$$X_i(k) = X_{i\text{Re}}(k) + j X_{i\text{Im}}(k) \quad (k=0, 1, \dots, N-1) \quad (1)$$

ここで、 i はフレーム番号、 j は虚数単位、 $X_{i\text{Re}}(k)$ は $X_i(k)$ の実部、 $X_{i\text{Im}}(k)$ は $X_i(k)$ の虚部をそれぞれ表す。

次に、最大振幅を有する周波数 k_{i0} を求める。一般に、振幅スペクトル値の大きい特徴的な成分に対して、そこからオクターブ離れた成分を操作してもその影響が知覚されにくいとされている[4]。そこで、 k_{i0} とオクターブの関係にある $1/2$ 倍あるいは 2 倍の周波数に対し、以下の埋め込み条件を満たすように埋め込み操作を施す。また、図1に提案法の埋め込みの概略図を示す。

- 透かし 1 の場合：ホスト信号の周波数 k_{i0} から 1 オクターブ離れた周波数 $k_{i0}/2$ と $2k_{i0}$ の位相の符号が逆
- 透かし 0 の場合：ホスト信号の周波数 k_{i0} から 1 オクターブ離れた周波数 $k_{i0}/2$ と $2k_{i0}$ の位相の符号が同じ

†中京大学工学部情報工学科, Chukyo University

‡近畿大学工学部情報学科, Kinki University

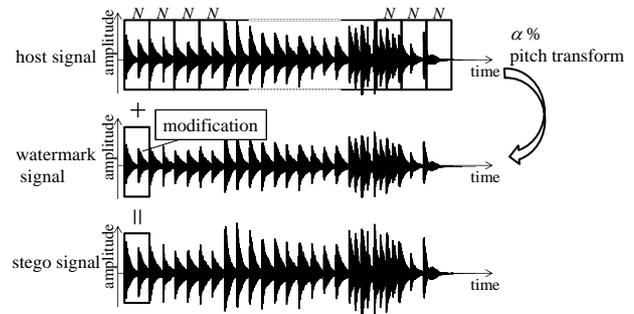


図1 提案法の埋め込みの概略図

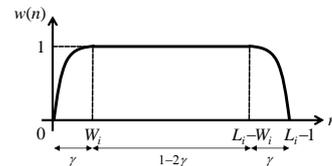


図2 窓関数 $w(n)$

上述の条件を満たしていない場合、図1のようにホスト信号を $\alpha\%$ ピッチ変換した透かし信号 $s(n)$ を、ホスト信号 $x(n)$ の値を用いて以下の式(2)にしたがって変更する。変更した後の透かし信号をホスト信号に加えることで、ホスト信号と逆位相の信号を得ることができる。

$$S'_i(k') = -X_i(k') \cdot \beta \quad (2)$$

ここで、 β は埋め込みの強度を表す変数、 k' は操作対象の周波数、 $S'_i(k')$ は変更した後の透かし信号の DFT 変換後の値である。ただし、周波数 $k_{i0}/2$ と $2k_{i0}$ のうち、ホスト信号と透かし信号の振幅スペクトル値の差が小さい周波数に対して式(2)の操作を行なう。変更した後の i 番目のフレームの透かし信号を i 番目のフレームに対応するホスト信号に加え、すべてのフレームに対して同様の処理を施すことでステゴ信号を得る。

本手法では、特定の周波数のみに埋め込み操作を施すのではなく、操作する周波数を選択できる利点がある。例えば、ホスト信号と透かし信号の周波数 $k_{i0}/2$ の振幅スペクトル値の差が大きい場合、式(2)による透かし信号の変更量は大きくなる。一方、ホスト信号と透かし信号の周波数 $2k_{i0}$ の振幅スペクトル値の差が小さい場合、周波数 $2k_{i0}$ のスペクトルを変更することで条件を満たすことができるため、特定の周波数のみに埋め込み操作を施す場合と比較して変更量を少なくでき、音質が向上すると考えられる。さらに、ホスト信号は操作していないので、ステゴ信号においてホスト信号の音質が維持されることから音質の向上が期待できる。

2.2 不連続性の回避

目的の値へ変更させた後、フレームの境界で信号の不連続が生じることがある。そこで、図2のような窓関数を

導入して不連続性を回避する。しかし、窓関数を導入することで位相が所望の値にならず、正しく透かしを埋め込むことができない場合がある。そこで、フレームの両端 γ は埋め込み対象外とし、 $1-2\gamma$ を埋め込み区間とする。これにより、 $1-2\gamma$ の部分は変化しないので透かしを正しく埋め込むことができる。しかし、 γ の部分を埋め込み対象外とした場合であっても、埋め込み範囲との境界で不連続になる。そこで、両端 γ の信号を、不連続が解消する方向へシフトさせ、この不連続性を回避する。最終的に窓関数を施すので、フレームの境界付近の透かし信号はゼロに近い値となる。そのため、信号をシフトさせた場合であっても音質の劣化は少ないと考えられる。

3. 実験

提案法の有効性を確認するために、音質および攻撃に対する耐性について文献[5]との比較実験を行なった。

3.1 実験条件

実験には、RWC 研究用音楽データベース(音楽ジャンル)から12曲を使用した。これらの楽曲は、16 bit 直線量子化、サンプリング周波数 44.1 kHz のステレオ音源である。埋め込みの評価基準は文献[6]にしたがい、埋め込み位置を検出するための同期符号として 63 bits の M 系列を使用した。

文献[5]の手法の埋め込みフレーム長は、文献[5]の実験で実際に使用されていた値である 1020 サンプルとした。提案法では、埋め込みフレーム長を $N=1020, 3675, 7238$ とした。これらの値は、文献[5]と同じフレーム長、文献[6]の埋め込み条件を満たすことができる最大のフレーム長、およびその中間の値として設定した。提案法に適した埋め込みフレーム長はステゴ信号の音質から検討する。また、 $\alpha=50, \beta=1.05, \gamma=0.1$ とし、 β は文献[5]と同程度の耐性をもつようにフレームごとに変更した。

3.2 実験結果

音質の客観的評価として、ホスト信号と非圧縮および MP3 圧縮を施した場合のステゴ信号との ODG を求めた。12 曲の平均値を表 1 に示す。また、音質の主観的評価の予備実験として、フレーム長を変更した場合の音質について被験者 1 名で調査した。その結果から、 $N=3675$ で 4.0 以上の音質が得られたため、主観的評価では $N=3675$ のステゴ信号を使用して実験した。被験者 10 名(男性 7 名、女性 3 名)による MOS の結果を図 3 に示す。表 1 および図 3 から、提案法の音質が文献[5]の手法より良いことを確認した。さらに、ホスト信号の音質より提案法のステゴ信号の音質のほうが良いと評価された楽曲が 6 曲あった。以上から、提案法の音質は良いと考えられる。

次に、文献[5]とフレーム長 $N=3675$ の提案法におけるステゴ信号に対して MP3 圧縮、MP3 圧縮 2 回符号化、および雑音付加後の透かしの BER を計算した。12 曲の BER の平均値を表 2 に示す。また、BER は以下の式(3)で求められる値である。

$$\text{BER} = \frac{\text{エラービット数}}{\text{30秒間に埋め込まれたビット数(180bits)}} \times 100[\%] \quad (3)$$

表 2 より、文献[5]では、埋め込み位置が正しく特定されず、BER が低い楽曲がみられた。一方、提案法では埋め込み位置を正しく特定でき、埋め込み位置が既知の場合の文献[5]と比較しても同程度の BER が得られたことを

表 1 ODG の平均値 : (a) 原曲とステゴ信号との比較, (b) 原曲とステゴ信号を MP3 圧縮した劣化信号との比較

	文献[5]	提案法		
		$N=1020$	$N=3675$	$N=7238$
(a)	-0.52	-0.54	-0.52	-0.53
(b)	-1.91	-1.63	-1.61	-1.60

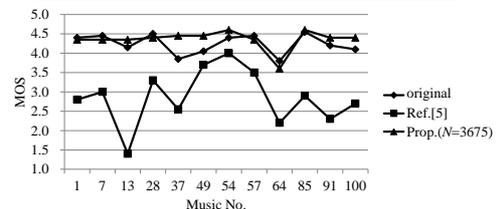


図 3 被験者 10 名による MOS の結果

表 2 BER [%] の平均値 : (a) MP3 128 kbps (joint stereo), (b) MP3 128 kbps (joint stereo) 2 回符号化, (c) ガウス性雑音付加 (overall average SNR 36 dB)

	BER [%]		
	文献[5]	既知	提案法
(a)	2.31	0.45	0.97
(b)	4.56	1.38	1.20
(c)	11.51	0.20	0.97

確認した。

4. まとめ

本稿では、ステゴ信号が音楽として成立する音質であれば、ホスト信号と異なる音が知覚された場合であっても問題はないという考えに基づき、ホスト信号をピッチ変換した音を透かし信号として埋め込む手法を提案した。既存の手法と比較した結果、攻撃に対する耐性および音質が良いことを確認した。

謝辞

本研究は JSPS 科研費 26870681, 26330214 の助成を受けたものです。

参考文献

- [1] 矢島尚幸, 大石邦夫, “振幅及び位相特性に基づく音楽信号への電子透かし,” 信学技報, SIP2004-134, pp.53-58, 2005.
- [2] H. Murata, A. Ogihara, M. Iwata and M. Yoshioka, “An audio watermarking method using sampled sound for polyphonic music,” Proc. 12th International Symposium on Communications and Information Technologies, pp.104-109, 2012.
- [3] Sound Engine Free, <http://soundengine.jp/software/soundengine/>, Accessed April 9, 2013.
- [4] 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, 2004.
- [5] W.N. Lie and L.C. Chang, “Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification,” IEEE Trans. on Multimedia, vol.8, no.1, pp.46-59, 2006.
- [6] “音響信号に対する電子透かし技術の評価基準,” <http://www.ieice.org/iss/emmm/ihc/audio/audio2013v2.pdf> (閲覧日: 2013 年 10 月 1 日).