

スペクトル拡散を用いたオーディオデータハイディング

1 V-5

大嶋良明 小林誠士 Dean Chen
日本アイ・ピー・エム 東京基礎研究所

はじめに

画像・ビデオ・音声などのデジタル・コンテンツに対する、著作権保護や改竄防止などへのニーズが高まっている。特にデジタル・オーディオにおいては、既存タイトルの新たな流通形態や、記録フォーマットの高品位化が提案されており、これらの傾向が顕著である。このような背景を考慮して、我々はCD品質のデジタルオーディオを対象とする電子透かし技術の開発を行っている。本稿では、スペクトル拡散に基づく手法を用いて、音楽コンテンツに対するメッセージ埋め込みの基礎実験を行った結果を報告する。また、埋め込み操作により生じた信号歪みを簡便に評価するため、MPEGオーディオのマスキングモデルの使用を検討したので、あわせて報告する。

メッセージの埋め込みと検出

オーディオ信号への情報埋め込み方法としては、DFTの位相項を変形する[1]、all-passフィルタを利用する[2]、エコー成分を付加する[1]、スペクトル拡散を利用する[1,3]などの手法が提案されている。これら手法にはそれぞれ長・短所があり、音質面での非可聴性と切り取り、再標本化、圧縮などに対するロバストネスを同時に実現するのは容易でない。ここではスペクトル拡散の一手法であるDSSS(Direct Sequence Spread Spectrum)を用いた。以下に離散時間システムにおける処理の概要を説明する。

まずbipolarな2値メッセージ系列 $x[n]$ を用意する。これを擬似ランダム系列、いわゆるchip系列 $p[n]$ により、広帯域化したもので正弦波搬送波 $s[n]$ を変調する。メッセージの埋め込みとは、ソースである音楽信号 $m[n]$ に対し、上記の方法でスペクトル拡散したメッセージを振幅調整して加える操作である。すなわち、

$$m'[n]=m[n]+a \cdot x[n] \cdot p[n] \cdot s[n]$$

となる。ここで a は搬送波振幅のスケーリング係数である。メッセージの検出は、 $m'[n]$ に再度chip系列を乗じ、unspreadした信号の搬送波成分を位相検出する事により実現する。ここで、

$$p^2[n]=1$$

であることにより、 $m'[n] \cdot p[n]$ の右辺は

$$m[n] \cdot p[n]+a \cdot x[n] \cdot s[n]$$

となる。 a の適切な設定により、第1項による位相誤差を軽減する。実験的に a を $m[n]$ の短時間におけるrms振幅値の2%程度とし、概ね良好な結果を得た。

基本性能の評価

埋め込み・検出の基本性能を観察するための実験を行った。まずメッセージの実効伝送レートを約30bpsと設定し、予備実験により搬送波周波数を検討した。このとき、搬送波が低域にある場合には音質劣化が顕著であったため、搬送波周波数を5.5kHzおよび11kHzとした。使用したコンテンツは、弦楽合奏による室内楽・ソプラノ独唱・ポピュラー音楽で、すべてCD音源を用いた。使用部分の演奏時間はそれぞれ、およそ1分35秒、50秒、4分10秒であった。

これらの試料に対し、chip周期を変化させて検出性能を測定した。図1にその結果を示す。いずれの試料においても、chip周期が搬送波の6ないし7周期分の場合に最適な結果が得られている。しかし、検出誤り率は試料により、1%から5%前後まで、ばらついており、これはコンテンツの性質の違いを反映したものと考えられる。

マスキングモデルによる音質評価

メッセージの埋め込み・検出において、埋め込みの強度が強いほど検出は容易になるが、その反面、音質劣化による聴感上の悪影響を生じる。これは、加法的雑音であり、原信号によるマスキング効果により評価できると考える。本稿ではマスキングの評価には、MPEGオーディオ規格[4]によるPsychoacoustic Model 1を使用した。以下に、処理の概要を述べる。

- 512点FFT分析の結果にもとづき、均等に帯域分割した32のサブバンド内での音楽信号のパワースペクトルを求める。
- スペクトルの局所的特徴をもとに、臨界帯域ごとに純音(tonal)成分と非純音(non-tonal)成分とを選別する。ここで可聴閾値以下の成分、および近傍成分でマスクされる純音成分はマスクーとしては除外する。
- 各臨界帯域において、周波数ごとに純音・非純音成分の個別マスキング閾値を求め、これらを可聴閾値に加味して、全体的マスキング閾値を求める。

以上で、マスキング現象により聴感上許容される歪みの目安が得られた。最後に透かし成分のスペクトルが、各周波数位置において、3で求めたマスキング閾値を越えないよう、振幅調整項をスケーリングする。

Data Hiding for Audio Using Spread Spectrum
Yoshiaki Ohshima, Seiji Kobayashi, and Dean Chen
IBM Research, Tokyo Research Laboratory
1623-14 Shimotsuruma, Yamato, Kanagawa 242,
Japan

図2に、その実施例を示す。破線に示すように拡散されたメッセージは、搬送波周波数(この例では5.5kHz)を中心として常に広帯域である。一方、音楽信号のスペクトルの特徴とその時間変化は、音源や楽曲の種類により様々である。この例の太線で示されるような、広帯域のスペクトルによるマスキングにおいては、効果的な振幅調整が可能となることがわかる。逆に、音楽信号のスペクトルが、狭帯域の場合や少数の鋭いピークを持つ場合には、得られるマスキング閾値カーブに深い谷部を生じるため、振幅調節は困難である。

まとめ

スペクトル拡散に基づく手法を用いて、CD品質の音楽コンテンツに対する埋め込み・検出の基本性能を調べた。また、MPEGオーディオの聴覚マスキングモデルを併用し、メッセージ埋め込みによる音質劣化の評価を試みた。

参考文献

1. N. Morimoto, "Techniques for Data Hiding in Audio Files", MIT Master Thesis, June 1995.
2. Y. Yardimci et. al., "Data Hiding in Speech Using Phase Coding", Eurospeech97, pp. 1679—1682, 1997.
3. L. Boney, et. al., "Digital Watermark for Audio Signals", Proc. IEEE MULTIMEDIA96, pp.473—480, 1996.
4. "Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s, Part 3:Audio", ISO/IEC 11172-3, 1993.

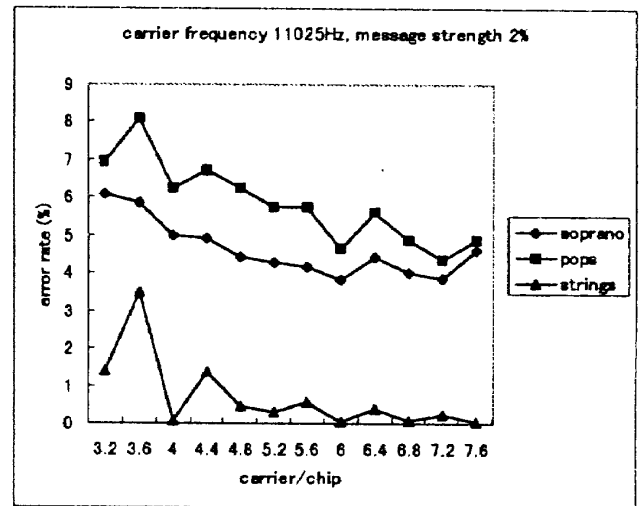


図 1

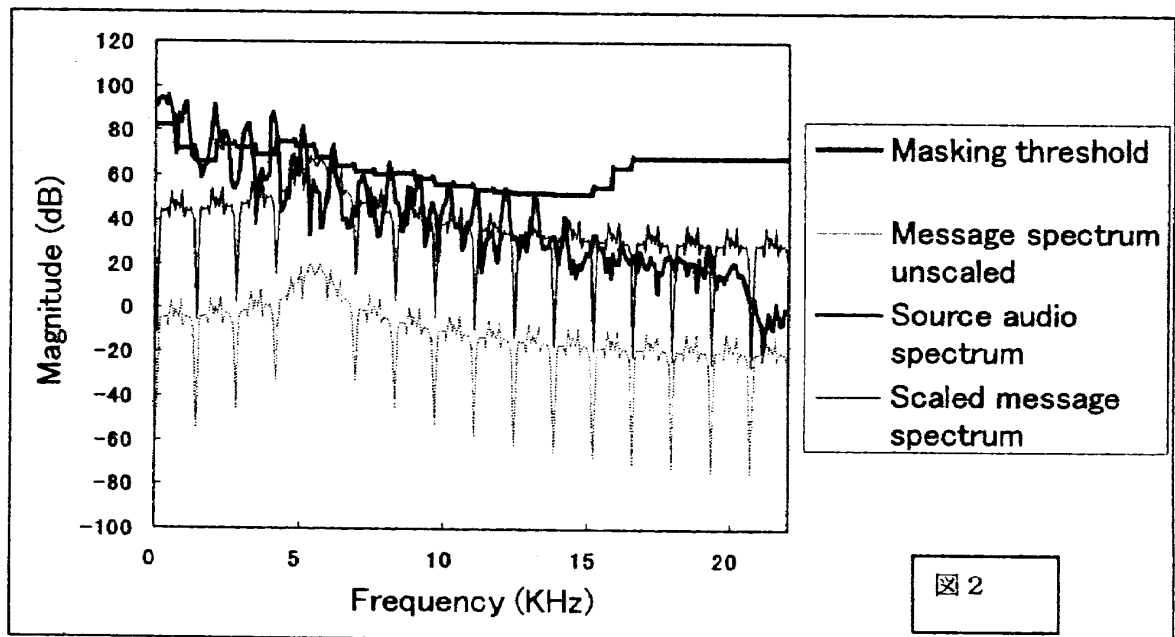


図 2