

非負値行列因子分解を用いた音楽電子透かしの埋め込み方法の考察

Study of Embedding Methods Using Nonnegative Matrix Factorization for Audio Watermarking

村田 晴美†
Harumi Murata

荻原 昭夫‡
Akio Ogihara

1. まえがき

音楽電子透かしに求められる要件のひとつとしてステゴデータの音質が劣化してはいけないことが挙げられる. この要件に対し, ホストデータとステゴデータとの間に差異があった場合でもステゴデータが音楽として成立していれば, 楽曲としての音質が劣化していないと定義する. そのために, モノラル信号の分離手法のひとつである非負値行列因子分解 (NMF) を利用した埋め込み法に着目する. 本稿では, NMF を用いてホストデータと協和音の関係になる音をステゴ信号として加えることで音楽上の違和感を低減させる 2 つの埋め込み法について, 音質と耐性の面から考察する.

2. 非負値行列因子分解を用いた音楽電子透かし

2.1 非負値行列因子分解

非負値行列因子分解 (NMF) とは, 非負値のデータを加法的な構成成分に分解することを目的とした多変量解析手法である [1]. NMF は, 時間周波数領域における振幅スペクトルで表された観測信号 $\mathbf{Y} (\in \mathbf{R}^{M \times N})$ を, 観測信号に含まれる各スペクトルパターンを有限の基底でモデル化した基底行列 $\mathbf{A} (\in \mathbf{R}^{M \times K})$ と, その基底が現れるタイミングとその大きさをモデル化したアクティベーション行列 $\mathbf{B} (\in \mathbf{R}^{K \times N})$ との積で近似表現する手法であり, 式(1)のように表される.

$$\mathbf{Y} \cong \mathbf{A}\mathbf{B} \quad (1)$$

また, ユークリッド距離規準の目的関数に対する乗算型更新式を式(2), (3)に示す.

$$a_{i,l} = \frac{[\mathbf{Y}\mathbf{B}^T]_{i,l}}{[\mathbf{A}\mathbf{B}\mathbf{B}^T]_{i,l}} a_{i,l} \quad (2)$$

$$b_{l,j} = \frac{[\mathbf{A}^T\mathbf{Y}]_{l,j}}{[\mathbf{A}^T\mathbf{A}\mathbf{B}]_{l,j}} b_{l,j} \quad (3)$$

ここで, $a_{i,l}, b_{l,j}$ は行列 \mathbf{A}, \mathbf{B} の要素である.

2.2 透かしの埋め込みおよび抽出

ホスト信号に対して窓幅が L サンプルで $L/2$ サンプルずつシフトさせる STFT を施す. 目的関数がユークリッド距離規準である NMF を用い, STFT により得られた振幅スペクトログラム \mathbf{Y} を基底行列 \mathbf{A} とアクティベーション行列 \mathbf{B} とに分解する. 図 1 にアクティベーション行列の例を示す. 分解されたアクティベーション行列 \mathbf{B} から各音高の発音時刻および消音時刻を推定する. 推定された発音時刻と消音時刻の間を 1 音の発音区間とし, 1 音に対して 1 ビットの透かしの埋め込む.

透かしの埋め込み前に, アクティベーション行列の係数

†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, Kindai University

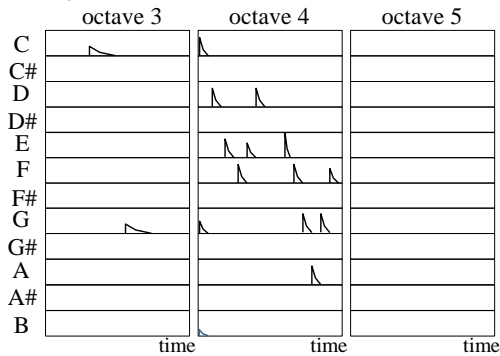


図 1 アクティベーション行列の例

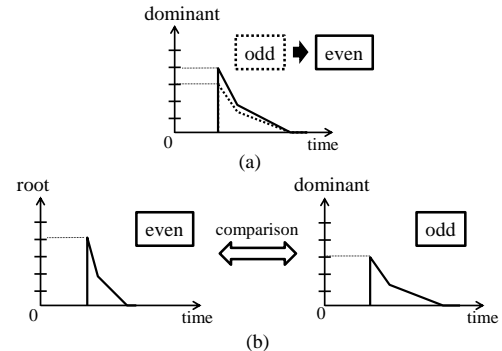


図 2 透かし 1 の埋め込み例: (a) 埋め込み規則 I の場合, (b) 埋め込み規則 II の場合

に対して量子化レベルを変更する. 量子化レベルを変更させることで, 攻撃に対して耐性をもたせることができると考えられる. 次に, 推定された音を根音とし, 根音に対して属音の関係にある音に対応する基底のアクティベーション行列の係数値を操作することで透かしの埋め込む. 本稿では, 透かしの埋め込み規則として以下の 2 パターン [2], [3] について考える. 例として, 図 1 のオクターブ 4 の C の音に透かし 1 を埋め込む場合について説明する. このとき, 属音はオクターブ 4 の G の音となる.

● 埋め込み規則 I

属音に対応するアクティベーション行列の係数の最大値の偶奇により透かしの埋め込む. 透かし 0 の状態を奇数, 透かし 1 の状態を偶数とする. 埋め込む透かしのビットと最大値の偶奇が等しい場合, 操作は何も行わず, 偶奇が異なる場合に属音のアクティベーション行列の係数値を図 2 (a) のように変更させて 1 ビットの透かしの埋め込む.

● 埋め込み規則 II

根音に対応するアクティベーション行列の係数の最大値と属音に対応するアクティベーション行列の係数の最大値の偶奇の関係により透かしの埋め込む. 根音と属音のアクティベーション行列の係数の最大

値がともに偶数、あるいはともに奇数の場合を透かし 0 の状態とし、偶奇が異なる場合を透かし 1 の状態とする。埋め込む透かしのビットと根音と属音の最大値の偶奇の関係が等しい場合、図 2 (b) のように操作は何も行わず、偶奇の関係が異なる場合に属音のアクティベーション行列の係数値を変更させて 1 ビットの透かしの埋め込む。

すべての音に埋め込み操作を行なった後、得られたアクティベーション行列 B' を用いて式(4)のように振幅スペクトログラムを再構成する。

$$Y' = AB' \quad (4)$$

振幅スペクトログラム Y' を用いて ISTFT を施し、ステゴ信号を得る。なお、各周波数の位相情報はホスト信号と同じものを用いる。

透かしの抽出は、埋め込みと同様にステゴ信号に対して窓幅が L サンプルで $L/2$ サンプルずつシフトさせる STFT を施す。得られた振幅スペクトログラム Y' に対して NMF を行ない、基底行列 A とアクティベーション行列 B' とに分解する。ここで、基底行列 A は埋め込みのときと同じ基底行列を使用する。これは、埋め込み時と抽出時の極端な分解の違いを低減させるためである。得られたアクティベーション行列 B' から各音高の発音区間を推定する。推定された音を根音とみなす。

埋め込み規則 I の場合、属音のアクティベーション行列の係数の最大値のみを用いて透かしの抽出する。最大値が奇数であれば透かし 0 を抽出し、そうでなければ透かし 1 を抽出する。埋め込み規則 II の場合、根音と属音のアクティベーション行列の係数の最大値の偶奇が同じであれば透かし 0 を抽出し、偶奇が異なる場合は透かし 1 を抽出する。この操作をすべての音に対して行ない、透かしの抽出する。

3. 実験

文献[4]から 10 曲を選択し、実験に使用した。これらの楽曲は簡単なピアノの練習曲であり、サンプリング周波数が 44.1 kHz、16 ビットに量子化されたステレオ音源の楽曲である。また、STFT のフレーム長を $L=8192$ 、埋め込み系列は文献[5]で定められているペイロードを使用し、左チャンネルのみに埋め込んだ。さらに、Cubase Artist7 でピアノの音源に対してオクターブ 3 からオクターブ 5 までの各種音階の音を作成し、作成した信号を教師信号として NMF を行ない、事前学習をした。事前学習によって得られた基底行列を埋め込みと抽出のときの基底行列 A として使用した。

3.2 音質評価実験

ステゴ信号の音質を評価するために、PQevalAudio による客観的音質の評価実験を行ない、ODG を算出した。

ODG 値の結果を表 1 に示す。評価基準[5]では、表 1(a) の最低値を -2.5 とし、表 1(b) については全楽曲の ODG の算術平均が -2.0 以上であることが求められている。表 1 の結果から、2 つの埋め込み手法ともに基準を満たすことができなかったが、音質は同程度であることがわかった。

3.1 攻撃に対する耐性実験

透かしの攻撃に対する耐性を評価するために、以下の 4 項目について評価をした。

- MP3 圧縮 128 kbps (joint stereo)
- MPEG4 HE-AAC 96 kbps

表 1 全楽曲の ODG 値の平均値：(a) 左チャンネルのみのホスト信号とステゴ信号との比較、(b) ホスト信号とステゴ信号を MP3 圧縮した劣化信号との比較

	ODG	
	(a)	(b)
埋め込み規則 I	-3.80	-3.32
埋め込み規則 II	-3.81	-3.36

表 2 全楽曲の BER の平均値：(a) MP3 圧縮、(b) AAC、(c) ガウス性雑音付加、(d) バンドパスフィルタ

	BER [%]			
	(a)	(b)	(c)	(d)
埋め込み規則 I	7.96	4.54	1.81	3.38
埋め込み規則 II	8.41	25.14	1.85	28.78

- ガウス性雑音付加 (overall average SNR 36 dB)
- バンドパスフィルタ 100 Hz–6 kHz, -12 dB/oct.

評価には、式(5)で算出される BER (bit error rate) を利用した。

$$BER = \frac{\text{エラービット数}}{\text{楽曲全体の埋め込みビット数}} \times 100[\%] \quad (5)$$

BER の結果を表 2 に示す。なお、本稿では 2 つの埋め込み法を比較するために、MP3 圧縮に対する耐性が文献[5]で定められている評価基準である 10 % 未満になるようにした。実験の結果、埋め込み規則 I のほうが攻撃に対する耐性が高いことがわかった。

4. まとめ

本稿では、非負値行列因子分解を利用した音楽電子透かし法の埋め込み法の考察を行なった。実験の結果、属音のアクティベーション行列の係数値のみで透かしの埋め込み手法のほうが良いということがわかった。

しかし、1 音に 1 ビットの透かししか埋め込んでいないため、埋め込み容量が十分であるとはいえない。そこで、埋め込み容量を増大させることが今後の課題として挙げられる。

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] 亀岡弘和, “非負値行列因子分解,” 計測と制御, vol.51, no.9, pp.835-844, 2012.
- [2] 村田晴美, 荻原昭夫, “楽曲の調判定と非負値行列因子分解を利用した音楽電子透かし法の検討,” 信学技報, vol.115, no.397, EMM2015-70, pp.55-60, 2016.
- [3] 村田晴美, 荻原昭夫, “非負値行列因子分解を用いた音楽電子透かし法の埋め込み容量の改善,” 信学技報, vol.116, no.34, EMM2016-15, pp.87-92, 2016.
- [4] “大人のためのやさしく弾けるクラシック名曲 100,” リットーミュージック, 2009.
- [5] “IHC Evaluation Criteria and Competition,” <http://www.ieice.org/iss/emm/ihc/IHC-criteriaVer4.pdf> (閲覧日: 2015 年 12 月 21 日).