

5R-5

伝達関数のスパース性仮定に基づく 音楽音響信号中のディレイエフェクトブラインド推定

阪上 大地[†] 安良岡 直希[‡] 高橋 徹[‡] 尾形 哲也[‡] 奥乃 博[‡]

[†] 京都大学 理学部理学科 [‡] 京都大学大学院 情報学研究科 知能情報学専攻

1. はじめに

本稿では音楽音響信号に含まれるディレイエフェクト(以下単にディレイと呼ぶ)のブラインド推定法について述べる. ディレイとは録音した楽器音や歌声を一定時間ごとに減衰させて元の信号に加える操作であり, コンサートホールの豊かな残響音をコンピュータ上で再現するリバーブエフェクトと並んで楽曲製作時に活用される主要なエフェクトの一つである. ディレイの周期や強度は製作時に固定され, 通常聴き手がこれらのパラメータを変更することはできない. しかし, ディレイのパラメータを推定して一旦除去を行い, パラメータを変更して再合成することができれば, ディレイの効果を聴き手の好きな条件に編集して聴くことが可能となる.

ディレイ推定操作の主たる課題は楽器の共鳴や残響の影響とディレイの影響を分解することである. ディレイの効果は自己回帰 (AR) フィルタで数的に記述できるが, 楽器の共鳴や残響も同じフィルタである程度記述できるので, フィルタ係数の推定だけではディレイのみの操作は困難である. 実際, 既存の残響抑圧法 [1] を用いて楽曲から付加したディレイ音の抑圧を試みたところ, ディレイ音の抑圧はできたが同時に楽器の共鳴音も抑圧され, 主にピアノ曲でスタッカート気味の演奏に変化するという問題が発生した.

本稿では, 特定の時間遅れ及びその整数倍に対応する部分のみに値を持つ AR フィルタを用いてディレイのみを推定する手法を報告する. 本手法は, まずディレイの時間間隔を推定し, 次にその間隔に基づき AR フィルタを設計し係数を推定する. 評価実験では本手法が楽器の共鳴音や残響音を歪ませることなくディレイのみを推定・抑圧可能であることを示す.

2. ディレイ推定手法

本稿では, ディレイの効果を記述するパラメータが未知であるという条件のもと, ディレイ音を含むモノラルの入力音響信号の時間周波数表現 $y_{n,l}$ からディレイ音を除いた音源信号 $s_{n,l}$ を復元する問題を扱う. ここで, n は時間フレーム, l は周波数ビンとする.

2.1 モデルの設計

一般的なディレイは以下の3要素で構成され, 図1のようなブロック線図で表される.

1. ディレイ周期 (D): ディレイ信号の繰り返し間隔.
2. 減衰率 (p): ディレイ音を繰り返す際の強度.
3. バランス (q): ディレイ信号の, 元の信号に対する相対強度. $q=1$ の時, 最初のディレイ信号は元の信号と同じ音量で再生される.

減衰率とバランスが同じ値の場合, ディレイは一つのフィードバック部をだけ持つ AR フィルタで表されると

Blind Music Delay Estimation and Separation Based on Sparseness of Transfer Function: Daichi Sakaue, Naoki Yasuraoka, Toru Takahashi, Tetsumiya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

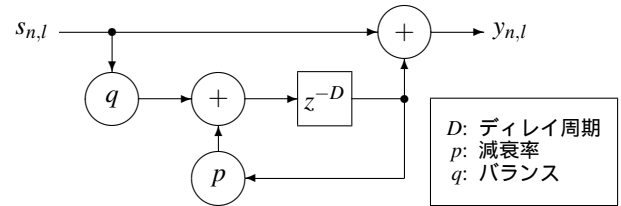


図1: ディレイの概念図

考えられる. また, 異なる値をとる場合でも, 特定の時間遅れ及びその整数倍の部分にフィードバックを持つ AR フィルタとなると考えられる. このようにディレイには AR フィルタ係数のスパース性が存在する.

上記ディレイの原理と既存の残響推定法 [2] を踏まえ, 以下のような統計モデルを立てる.

1. 音源信号 $s_{n,l}$ は各時間フレーム, 各周波数ビン上で平均 0, 分散 $S_{n,l}$ の複素正規分布に従う.
2. 観測信号 $y_{n,l}$ は音源信号 $s_{n,l}$ によって駆動される有限長の AR システムによって生成される. ただし, このシステムは, ディレイ周期を基本とした疎なパラメータによって記述されている. ディレイ周期を D , 時間フレーム幅を W とすると, ディレイを説明する AR フィルタ係数は D の整数倍の位置を中心とした幅 $2W+1$ の範囲にのみ存在している. 計算は一定のフィルタ次数 K で打ち切る.

$$y_{n,l} = \sum_m \sum_{mD-W \leq k \leq mD+W} g_{k,l} y_{n-k,l} + s_{n,l} \quad (1)$$

$g_{k,l}$ は第 l 周波数ビンの第 k AR フィルタ係数である. これ以降 AR フィルタ係数の存在するフレーム番号の全体を Λ で表す.

このモデルのパラメータは $\{g_{k,l}, S_{n,l}, D, W\}$ である.

2.2 パラメータ推定

本手法のパラメータ推定アルゴリズムを表1に示す. なお, ディレイの時間フレーム幅 W については実験的に最適値 5 を定めた.

2.2.1 ディレイ周期推定

本手法では, ディレイを記述するパラメータのうち, ディレイ周期 D を従来法に基づきまず最初に推定する. 残響抑圧法 [2] によって推定される AR フィルタ係数 $\tilde{g}_{k,l}$ にはディレイ周期に応じた鋭いピークが存在するため,

表1: パラメータ推定アルゴリズム.

1. 式 (2) に基づきディレイ周期 D を推定
2. AR フィルタ係数 $g_{k,l}$ を 0 で初期化, 分散 $S_{n,l}$ を $S_{n,l} = |y_{n,l} - \sum_{k \in \Lambda} g_{k,l} y_{n-k,l}|^2$ で初期化
3. AR フィルタ係数 $g_{k,l}$ 及び分散 $S_{n,l}$ を推定
 - 3.a 式 (5) に基づきフィルタ係数を更新する
 - 3.b 式 (6) に基づき分散を更新する
 - 3.c パラメータ変化が十分小さくなるまで 3.a から反復

表 2: 実験条件 .

サンプリングレート	44.1kHz
STFT 窓	1024 点 Gaussian
STFT シフト幅	256
フィルタ次数 (K)	323

表 3: 推定結果: 左から順に入力信号, 残響抑圧信号, ディレイ抑圧信号の対数スペクトル距離 LSD (dB).

		成功時		失敗時			
バランス		成功時		失敗時			
Pops	0dB	6.041	4.607	4.502	5.631	5.065	5.755
	-12dB	2.703	2.102	1.296	—	—	—
	-24dB	0.948	1.866	0.745	—	—	—
Jazz	0dB	3.167	2.706	2.494	—	—	—
	-12dB	1.731	2.404	1.267	1.545	2.244	1.945
	-24dB	—	—	—	0.497	2.197	1.855

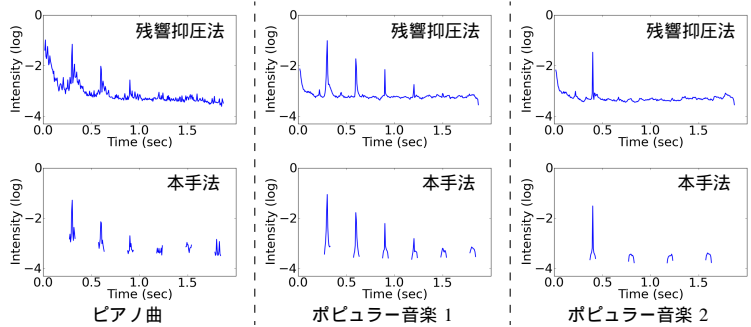


表 4: 種類の楽曲に対する AR フィルタ係数の例. 上段は残響抑圧法により推定された AR フィルタ, 下段は本手法に基づくスパースな AR フィルタ. ポピュラー音楽 2 は減衰率とバランスが一致している例.

AR フィルタ係数のパワー周波数平均 $\tilde{G}_k = \sum_{l=0}^{L-1} |\tilde{g}_{k,l}|^2$ の二次差分の最小値を与えるフレーム番号 k をディレイ周期推定値とする. ただし, AR フィルタの低次係数では残響の効果が支配的であるため, 周期推定では 10 フレーム以下を無視する.

$$D = \underset{k}{\operatorname{argmin}} (\tilde{G}_{k-1} + \tilde{G}_{k+1} - 2\tilde{G}_k) \quad (2)$$

2.2.2 AR フィルタ係数の推定

前述の統計モデルによれば, 観測信号全体 $\mathbb{Y} = \{y_{n,l}\}_{0 \leq n \leq N-1, 0 \leq l \leq L-1}$ が与えられたときのパラメータ $\{S_{n,l}\}_{0 \leq n \leq N-1, 0 \leq l \leq L-1}, \{g_{k,l}\}_{k \in \Lambda, 0 \leq l \leq L-1}$ の最尤推定は, 以下のコスト関数 Q の最小化問題となる.

$$Q = \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} \left(\log S_{n,l} + \frac{|y_{n,l} - \sum_{k \in \Lambda} g_{k,l} y_{n-k,l}|^2}{S_{n,l}} \right) \quad (3)$$

この最尤推定は, $S_{n,l}, g_{k,l}$ を交互に繰り返し推定することで実現できる. 上式の AR フィルタ $g_{k,l}$ に対する偏微分を 0 と置き連立方程式を解くと, $g_{k,l}$ に対する更新式が以下のように導出できる.

$$r_{i,j}^{(l)} = \sum_{n=0}^{N-1} \frac{y_{n-i,l}}{S_{n,l}} y_{n-j,l}^* \quad (4)$$

$$\begin{bmatrix} g_{k_1,l} \\ \vdots \\ g_{k_I,l} \end{bmatrix} = \begin{bmatrix} r_{k_1,k_1}^{(l)} & \cdots & r_{k_I,k_1}^{(l)} \\ \vdots & \ddots & \vdots \\ r_{k_1,k_I}^{(l)} & \cdots & r_{k_I,k_I}^{(l)} \end{bmatrix}^{-1} \begin{bmatrix} r_{0,k_1}^{(l)} \\ \vdots \\ r_{0,k_I}^{(l)} \end{bmatrix} \quad (5)$$

ただし, I は Λ の要素数, $\Lambda = \{k_1, \dots, k_I\}$ とする. この更新式は残響抑圧法 [2] の更新式と類似しており, $k \in \Lambda$ に対応する行及び列のみが式中に現れる点が異なる. 最後に, 得られた $g_{k,l}$ の値を用いて $S_{n,l}$ の値を更新する.

$$S_{n,l} = \left| y_{n,l} - \sum_{k \in \Lambda} g_{k,l} y_{n-k,l} \right|^2 \quad (6)$$

3. 評価実験

本ディレイ推定法の性能を評価するため, ディレイを付加した音響信号からディレイ成分を推定・抑圧する実験を行った.

3.1 実験条件

RWC 音楽データベース [3] に含まれる楽曲のうち, ポピュラー音楽, No.1-3 の 3 楽曲及び, ジャズ, No.1-3 のピアノ曲 3 楽曲を用いる. これらの楽曲に対しインパルス応答によって記述したディレイ音を畳み込み, ブラインド推定を行う. ディレイ周期は 200ms から 500ms,

周期ごとの減衰率は 0dB から -96dB, バランスは 0dB から -24dB とする. その他の実験条件は表 2 に付記する.

ディレイ推定の性能は 1) ディレイ周期が正しく推定できたか, 2) ディレイ抑圧後の信号がディレイ付加前の信号にどれだけ近いのか, との 2 点で評価する. そこで, まずディレイ周期 D が真値から誤差 5 フレームに収まっているかどうかで周期推定の成功・失敗を判断する. 次にディレイ抑圧結果と真の音源信号の間の対数スペクトル距離 (LSD) を算出する. LSD は次式で定義される.

$$\text{LSD}(\mathbb{S}, \hat{\mathbb{S}}) = \sqrt{\sum_{n=0}^{N-1} \sum_{l=0}^{L-1} \left(20 \log_{10} \left| \frac{S_{n,l}}{\hat{S}_{n,l}} \right| \right)^2} / NL \quad (7)$$

ここで, $\mathbb{S} = \{S_{n,l}\}_{0 \leq n \leq N-1, 0 \leq l \leq L-1}$ は真の音源信号, $\hat{\mathbb{S}} = \{\hat{S}_{n,l}\}_{0 \leq n \leq N-1, 0 \leq l \leq L-1}$ は推定された音源信号である.

3.2 実験結果

バランスが -12dB 以上である場合, ポピュラー音楽, ジャズ音楽のほぼ全てのデータでディレイ周期推定に成功した. ポピュラー音楽の場合, 実験を行った最小値である -24dB までほぼ推定できていたが, 最初のディレイ信号の強度が 0dB でも推定に失敗することがあった. これは式 (2) によるディレイ周期推定が完全ではなく, 推定に失敗したためである.

本稿の手法により得られた対数スペクトル距離を表 3 に記す. ディレイ周期推定に成功している場合はディレイの効果を選択的に打ち消し, スパース性を仮定しない残響抑圧法に比べ AR フィルタの性能が改善されていることが分かる. 表 4 に推定されたスパースな AR フィルタの例を示す.

4. おわりに

本稿では, AR フィルタ係数のスパース性仮定に基づくディレイ音の選択的抑圧法を報告した. 本研究の今後の課題としては, ディレイ周期推定の高精度化, 複数のディレイが加えられた音楽音響信号に対する推定などが考えられる. なお, 本研究は科研費 (S), グローバル COE, CrestMuse の支援を受けた.

参考文献

- [1] N. Yasuraoka *et al.*: Music dereverberation using harmonic structure source model and Wiener Filter, *ICASSP-2010*, 53-56.
- [2] T. Yoshioka *et al.*: An integrated method for blind separation and dereverberation of convolutive audio mixtures, *EUSIPCO-2008*.
- [3] 後藤他: RWC 研究用音楽データベース, 情報論, Vol.45, No.3 (2004), 728-738.