調波GMMとWienerフィルタに基づく音楽音響信号の残響抑圧

安良岡 直希† 吉岡 拓也‡† 中谷 智広‡ 中村 篤‡ 奥乃 博†
†京都大学 大学院情報学研究科 知能情報学専攻 ‡NTTコミュニケーション科学基礎研究所

1. はじめに

本稿では音楽音響信号に対する残響抑圧について述べる. 残響は音楽の演奏や鑑賞の場面では楽曲の印象を豊かにすべく積極的に活用されている半面, 楽曲解析のための音響信号処理の場面では弊害となることが多い. 残響の影響を取り除く事ができれば, 楽曲検索などのための自動採譜や音源分離の精度向上が期待される. また, 既存楽曲の残響効果を自由に操作できるようになれば個人の嗜好に合わせた音楽鑑賞システムへと応用できる.

音楽音響信号での残響の特徴は, コンサートホールで見られるように, 強度が強く時間が長いことである. 残響抑圧は従来より音声信号を対象に広く研究されているが [1], 音楽音響信号は想定されてこなかった. 我々が実際に既存の残響抑圧法 [2, 3] を音楽信号に適用したところ, 二つの問題点：1) 全極モデルやノンパラメトリックモデルによる音源スペクトル表現が音楽音響信号の特徴である調波構造と対応しにくい, 2) 逆畳み込みによる残響抑圧処理では強くて長い残響を取りきることはできない, が発見された.

本稿では音楽音響信号に適した残響抑圧法として, 1) 調波 Gaussian Mixture Model (GMM) [4] による音源スペクトル推定, 2) 音源モデル情報を利用した Wiener フィルタによる残響抑圧処理, に基づく新たな手法を報告する. Wiener フィルタは従来の線形フィルタでは取りきれない残響成分を抑圧するのに有効である. また, 調波 GMM による音源スペクトル情報を元に Wiener フィルタを設計することで, 音源成分を高精度に抽出する.

2. 本手法における残響抑圧処理

本稿では, 残響特性が未知のもと, モノラル音響信号の時間周波数表現 $y_{n,l}$ から音源信号 $s_{n,l}$ 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.

本残響抑圧法は, 従来法 [2] と同じく, 以下の仮定からなる観測モデルに基づき定式化される.

1. 音源信号 $s_{n,l}$ は各時間フレーム, 各周波数ビンにて平均 0, 分散 $\lambda_{n,l}$ の複素正規分布に従う. このとき $\lambda_{n,l}$ はパワースペクトル密度 (PSD) と対応する.
2. 観測信号 $y_{n,l}$ は音源信号によって駆動される K 次の自己回帰システムによって生成される.

$$y_{n,l} = \sum_{k=1}^{K} g^*_{k,l} y_{n-k,l} + s_{n,l} \quad (1)$$

ここで, $*$ は複素共役を表し, $g_{k,l}$ は第 l 周波数ビンの第 k 自己回帰係数である. $g_{k,l}$ を本稿では残響フィルタと呼ぶ.

観測信号全体 $\mathbb{Y} = \{y_{n,l}\}_{0\le n\le N-1, 0\le l\le L-1}$ が与えられたときのパラメータ $\{\lambda_{n,l}\}_{0\le n\le N-1,0\le l\le L-1}$, $\{g_{k,l}\}_{1\le k\le K,0\le l\le L-1}$ の最尤推定は, 以下に示す板倉斎藤擬距離と等価なコスト Q の最小化問題となる.

$$Q = \sum_{n=0}^{N-1}\sum_{l=0}^{L-1}\left(\log\lambda_{n,l} + \frac{|y_{n,l} - \sum_{k=1}^{K} g^*_{k,l} y_{n-k,l}|^2}{\lambda_{n,l}}\right) \quad (2)$$

この最尤推定は, $\lambda_{n,l}$, $g_{k,l}$ を交互に繰返し推定することで達成される. なお, 残響フィルタ $g_{k,l}$ の更新方法は従来法 [2] をそのまま用いる.

2.1 調波 GMM 音源モデルに基づく音源推定

音源生成過程における音源 PSD $\lambda_{n,l}$ は [3] のようにそのままノンパラメトリックに扱うことも可能だが, 対象となる音源の性質を良く表すようにモデル化する方が高い残響抑圧性能が得られると期待される. 吉岡らは音声を対象に, フォルマント構造を表現する全極モデルにより $\lambda_{n,l}$ をモデル化した [2]. しかしこれは音楽音響信号において重要な要素である調波構造を明示的に表現しない.

我々は残響抑圧法を音楽音響信号に適用するため, 基本周波数推定 [4] 及び楽器音の分析合成 [5] のためのモデルとして使われている調波 GMM を用いる. 音源 PSD が J 個の調波構造と一つの残差成分の重み付き混合からなると仮定し, それぞれを混合ガウス分布でモデル化する.

$$\lambda_{n,l} = \sum_{j=1}^{J}\left(w_n^{(H)}(j)\sum_{m=1}^{M} H_{n,l}(j,m)\right) + w_n^{(I)}\sum_{i=1}^{I} I_{n,l}(i) \quad (3)$$

$$H_{n,l}(j,m) = \frac{u_n(j,m)}{\sqrt{2\pi\sigma^2}}\exp\left[\frac{-(\omega_l - m\mu_n(j))^2}{2\sigma^2}\right] \quad (4)$$

$$I_{n,l}(i) = \frac{v_n(i)}{\sqrt{2\pi\gamma^2}}\exp\left[\frac{-(\omega_l - \nu(i))^2}{2\gamma^2}\right] \quad (5)$$

ここで, ω_l は周波数ビンから Hz への写像であり, $w_n^{(H)}(j)$, $\mu_n(j)$, $u_n(j,m)$, σ^2 はそれぞれ第 j 調波構造の重み, 基本周波数, 第 m 倍音相対強度, 調波ピークの広がりを表す分散である. 一つのガウス分布が一つの調波ピークに対応する. 残差に対応させる式 (3) 第 2 項も GMM であるが, 予め設定した固定の平均 $\nu(i)$ と広い分散 γ^2 を用い, 重み $w_n^{(I)}$ 及び相対強度 $v_n(i)$ だけを適応させる. この残差成分用モデルは, 調波構造モデルの谷にパワーが全くないと仮定されることで残響フィルタ推定が不安定になるのを防ぐのに重要である.

調波 GMM 音源モデルの推定は式 (2) で示したとおり板倉斉藤擬距離の最小化問題であるが, 調波 GMM のパラメータ $\{w_n^{(H)}(j), u_n(j,m), \sigma^2, \mu_n(j), w_n^{(I)}, v_n(i)\}$ を板倉斉藤擬距離で解析的に更新することは困難であるため, 代わりに Kullback-Leibler (KL) 擬距離 Q' の制約付き最小化問題に近似する.

$$Q' = \sum_{n=0}^{N-1}\sum_{l=0}^{L-1}|\bar{s}_{n,l}|^2\log\frac{|\bar{s}_{n,l}|^2}{\lambda_{n,l}}. \quad (6)$$

ただし, $\forall n:\ \sum_{j=1}^{J} w_n^{(H)}(j) + w_n^{(I)} = \sum_{l=0}^{L-1}|\bar{s}_{n,l}|^2$,
$\forall j,n: \sum_{m=1}^{M} u_n(j,m) = 1, \forall n: \sum_{i=1}^{I} v_n(i) = 1$

Music Dereverberation Using Harmonic Structure Source Model and Wiener Filter: *Naoki Yasuraoka (Kyoto Univ.), Takuya Yoshioka (NTT, Kyoto Univ.), Tomohiro Nakatani, Atsushi Nakamura (NTT), and Hiroshi G. Okuno (Kyoto Univ.)*

表 1: 実験条件.

サンプリングレート	44.1 kHz
STFT 窓	1024 pt Gaussian
STFT シフト幅	256
J : 調波 GMM の仮定音源数	8
P : 全極モデルの次数	128

表 2: 残響抑圧結果: hgmm=調波 GMM, ap=全極モデル, np=ノンパラメトリックモデル, deconv=逆畳み込み, wiener=Wiener フィルタ.)

Method	β	α	LSDI(dB)
hgmm_deconv	-	-	**0.938**
ap_deconv	-	-	0.628
np_deconv	-	-	0.367
hgmm_wiener	2.0	0.8	**1.511**
ap_wiener	2.5	0.0	1.308
np_wiener	2.5	1.0	1.156

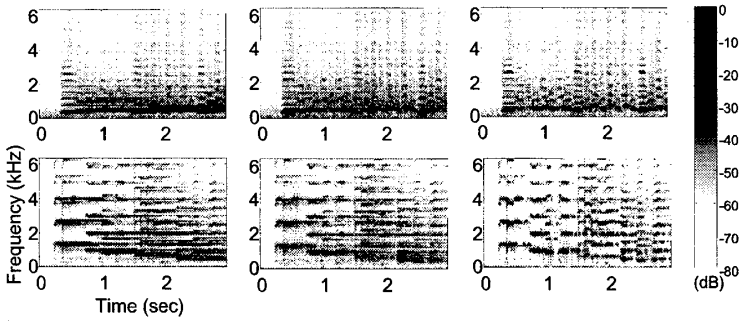

図 1: 残響信号とその抑圧結果: (上) フルート, (下) バイオリン, それぞれ左から順に観測信号, 全極モデル・逆畳み込みによる残響抑圧結果, 調波 GMM・Wiener フィルタによる残響抑圧結果.

ここで, $\bar{s}_{n,l} = \left| y_{n,l} - \sum_{k=1}^{K} g^*_{k,l} y_{n-k,l} \right|^2$ は反復推定途中の残響フィルタを用いて推定した音源信号である. 板倉斉藤擬距離と KL 擬距離は等価ではないが, パラメータが発散するような結果にはならないことを実験的に確認した. 調波 GMM の KL 擬距離による最小化は EM アルゴリズムを用いて効率的に解くことができる. 例えば基本周波数 $\mu_n(j)$ は M ステップにおいて以下の式で更新される.

$$\mu_n(j) = \frac{\sum_{m=1}^{M} \sum_{l=0}^{L-1} m\omega_l \bar{H}_{n,l}(j,m)}{\sum_{m=1}^{M} \sum_{l=0}^{L-1} m^2 \bar{H}_{n,l}(j,m)} \tag{7}$$

ここで, $\bar{H}_{n,l}(j,m)$ は E ステップで求められる第 j 番目調波 GMM の m 次高調波成分の期待値である.

2.2 Wiener フィルタを用いた残響抑圧

音楽音響信号中の長くて強い残響を抑圧するために, 本手法では Wiener フィルタを用いる. Wiener フィルタに基づく音源信号の推定値 $\tilde{s}_{n,l}$ は以下の式より導かれる.

$$\tilde{s}_{n,l} = W_{n,l} y_{n,l} \tag{8}$$

$W_{n,l}$ は Wiener ゲインであり,

$$W_{n,l} = \frac{\kappa_{n,l}}{\kappa_{n,l} + \alpha |r_{n,l}|^2} \tag{9}$$

$$\kappa_{n,l} = \beta \hat{\lambda}_{n,l} + (1-\beta) |\hat{s}_{n,l}|^2 \tag{10}$$

と定義する. ここで $r_{n,l} = \sum_{k=1}^{K} \hat{g}^*_{k,l} y_{n-k,l}$ は推定された残響フィルタ $\hat{g}^*_{k,l}$ により算出される残響成分である. α (>0) は残響の抑圧強度を調整するパラメータであり, β ($0 \le \beta \le 1$) は推定後の音源モデルが示すパワースペクトル $\hat{\lambda}_{n,l}$ と逆畳み込みによる残響抑圧結果が示すパワースペクトル $|\hat{s}_{n,l}|^2 = |y_{n,l} - r_{n,l}|^2$ との混合比である. $\hat{\lambda}_{n,l}$ がうまく推定されていれば, $\beta > 0$ とすることで $\hat{s}_{n,l}$ を直接出力とする従来法に比べて線形フィルタで取りきれない残響成分も抑圧できると期待される

3. 評価実験

本手法の音楽音響信号に対する残響抑圧性能を評価するために行ったシミュレーション実験について述べる. MIDI 音源によって合成した残響の存在しない音響信号を真の音源信号とし, これに残響時間 RT_{60} が 1 秒を越えるインパルス応答を畳み込んだものを観測信号として残響抑圧を行う. 評価データには無伴奏演奏を用い, バイオリン 3 曲, フルート 3 曲, チェロ 3 曲の計 9 曲を用いる. 残響抑圧の精度は抑圧結果と音源信号との間の対数スペクトル距離改善量 (LSDI) :

$$\mathrm{LSDI} = \mathrm{LSD}(\mathbb{Y}, \mathbb{S}) - \mathrm{LSD}(\hat{\mathbb{S}}, \mathbb{S}) \tag{11}$$

$$\mathrm{LSD}(\eta, \xi) = \sqrt{\sum_{n=0}^{N-1} \sum_{l=0}^{L-1} (20 \log_{10} \left| \frac{\eta_{n,l}}{\xi_{n,l}} \right|)^2 / NL} \tag{12}$$

によって評価する. ここで, $\mathbb{S} = \{s_{n,l}\}_{0 \le n \le N-1, 0 \le l \le L-1}$ は真の音源信号, $\hat{\mathbb{S}} = \{\hat{s}_{n,l}\}_{0 \le n \le N-1, 0 \le l \le L-1}$ は残響抑圧後の信号である. その他の実験条件を表 1 に記す.

本稿で報告した調波 GMM と Wiener フィルタの組み合わせと, 比較対象の従来法として全極モデルおよびノンパラメトリックモデル, 逆畳み込み法を用いた残響抑圧結果に対する LSDI を表 2 に記す. なお Wiener フィルタパラメータ α と β は最適値を実験的に定めた. 音源モデルの比較では調波 GMM モデルを用いた本手法が, 残響抑圧処理方法では Wiener フィルタによるものがそれぞれより良い結果となっている. また, 最適な音源モデル比 β が全極モデルでは 0 に近いのに対し, 調波 GMM 及びノンパラメトリックモデルでは 1 に近く, 調波 GMM 音源モデルが音響信号をより的確にとらえることで Wiener フィルタ精度を向上させていることが分かる.

図 1 に実演奏音響信号に対する残響抑圧結果のスペクトログラムを示す. 右端の本手法を用いたものが最も単音が明瞭に確認できることから, 残響成分が確実に抑圧されていることが分かる.

4. おわりに

本稿では, 調波構造音源モデルと Wiener フィルタを用いた音楽音響信号向けの新しい残響抑圧法を報告した. 今後は, 本手法を多重奏音響信号へ適応した場合の挙動等について評価していく. なお, 本研究は, 科研費, GCOE の支援を受けた.

参考文献

[1] Gillespie et. al. Strategies for improving audible quality and speech recognition accuracy of reverberant speech. In *Proc. ICASSP*, Vol. 1, pp. 676–679, 2003.
[2] Yoshioka, et. al. An integrated method for blind separation and dereverberation of convolutive audio mixtures. In *Proc. EUSIPCO*, 2008.
[3] Nakatani, et. al. Blind speech dereverberation with multi-channel linear prediction based on short time Fourier transform representation. In *Proc. ICASSP*, pp. 85–88, 2008.
[4] Kameoka, et. al. Extraction of multiple fundamental frequencies from polyphonic music using harmonic clustering. In *Proc. ICA*, pp. I–59–62, 2004.
[5] 安部他. 音色の音高依存性を考慮した楽器音の音高操作手法. 情処学論, Vol. 50, No. 3, pp. 1054–1066, 2009.