調波成分分析による音楽信号の劣決定ブラインド音源分離*

◎細谷弘 (東大・工), 和泉洋介, 小野順貴, 嵯峨山茂樹 (東大・情報理工)

1 はじめに

音楽信号を対象とした音源分離は, 音楽検索, 加工, 自動採譜といった様々な応用の基礎技術として, 重要性が認識されつつある. 信号源分離問題全般に広く用いられている手法としては独立成分分析 [1] があるが, 音楽信号はステレオ録音された対象が多く, 音源数がマイクロフォン数よりも多い劣決定な状況がしばしば発生するため, 独立成分分析が容易に適用できない場合も多く存在する. こうした状況下で音源分離を行うためには, 音源信号にさらに強い仮定を設けた理論が必要であり, 従来研究においては, 音源信号のスパース性を利用したものが多く行われている [2]. これは音声信号を対象にする場合では有効な性質であるが, 音楽信号を対象とする場合では, 楽曲構成の性質上, 必ずしも有効ではない.

本稿では, 音楽信号の重要な性質である調波性を積極的に活用し, 音楽信号に特化した新たな音源分離手法, 調波成分分析を提案する.

2 本研究の着眼点

2.1 音楽信号の非スパース性

スパース性とは信号のエネルギーがまばらにしか存在しない性質のことであり, 音声分離の研究で多く利用されている.

これは音楽信号においてもある程度成り立つと思われるが, 互いの倍音同士が重なることが多いため, 互いの成分が独立ではなく, スパース性による分離のアプローチは必ずしもうまく行かないと考えられる.

2.2 音楽信号の調波性の利用

本研究のアプローチは, 音楽信号の調波性を利用することである. 調波性を持つ音響信号は, 基本周波数成分と倍音成分にエネルギーが集中しているので, それらから大きく外れた周波数成分では, 観測信号に対する寄与がほとんどないと考えられる. 従って, 各音源信号の各時刻での基本周波数が既知であれば, 個々の時間周波数成分において, 寄与する音源数をもとの音源数よりも少なく考えることができるため, 分離が容易になると予想できる.

3 調波成分分析の定式化

3.1 観測モデル

信号は全て時間周波数領域で表現し, 各時間周波数 bin で, j 番目の音源信号, 観測信号を j 番目の要素とする複素ベクトル $\mathbf{S}_{\omega,t}$, $\mathbf{O}_{\omega,t}$ を定義する. ここで,ω と t はそれぞれ, 周波数と時間フレームのインデックスである. 混合行列として複素行列 A_ω を定義し, 観測モデルを

$$\mathbf{O}_{\omega,t} = A_\omega \mathbf{S}_{\omega,t} \tag{1}$$

とし, 畳み込み混合の形で定義する. ここでは A_ω の各列ベクトルの二乗ノルムは 1 に規格化されていると仮定し, 一般的にブラインド信号処理で生じるスケールの任意性を解消するものとする.

3.2 音楽信号の調波性に基づく定式化

簡単のために, 各音源が各時刻で調波性を持ち, 単音の信号であるような楽曲を対象とする. 従って, 時刻 t の j 番目の音源信号 $S_{j,\omega,t}$ は周期 $T_{j,t}$ の周期信号であるとみなせる.

調波成分分析の目的は, 独立成分分析に置ける独立性の代わりに, 調波性を表す何らかの $S_{j,\omega,t}$ の関数 (これはピッチ周期 $T_{j,t}$ も変数として含む) を定義して, (1) 式を満たすような $S_{j,\omega,t}$, A_ω を求めることである. 調波性の定義の仕方はいろいろありえるが, ここでは周期性の定義が $f(t) = f(t+T)$ であることに着目し, $\int |s(t) - s(t+T)|^2 dt$ を周期性のコストと考える。これを Parseval の等式を用いて周波数領域で表すと,

$$\begin{aligned}\int |s(t)-s(t+T)|^2 dt &= \frac{1}{2\pi}\int |\mathcal{F}[s(t)-s(t+T)]|^2 d\omega \\ &= \frac{2}{\pi}\int \sin^2\frac{\omega T}{2}|S(\omega)|^2 d\omega \end{aligned} \tag{2}$$

と導かれ, これは, $S_{j,\omega,t}$ に

$$q(\omega, T_{j,t}) = \sin\frac{\omega T_{j,t}}{2} \tag{3}$$

で定義される櫛形ノッチフィルタ $q(\omega, T_{j,t})$ を乗じて生じる, 非調波誤差 $P_{j,\omega,t} = q(\omega, T_{j,t}) S_{j,\omega,t}$ の二乗積分を考えることと同値である. 本研究ではまず, これを調波性の指標として利用した. $P_{j,\omega,t}$ が平均 0, 分散 σ_j^2 のガウス分布に従い, 互いに独立と仮定すると, 対数尤度 $\log p(\mathbf{P}) = \sum_{j,\omega,t} \log p(P_{j,\omega,t})$ が定義できる. これを (1) 式のもとで最大化することで, 各 $P_{j,\omega,t}$ が最小となる $\mathbf{S}_{\omega,t}$ を求めることができる. 目的関数の具体形は

$$\begin{aligned} J = &-mn\sum_j \log\sigma_j^2 - \sum_{\omega,t} \mathbf{S}_{\omega,t}^H Q_{\omega,t}^2 \mathbf{S}_{\omega,t} \\ &- \sum_{\omega,t} \boldsymbol{\lambda}_{\omega,t}^H (\mathbf{O}_{\omega,t} - A_\omega \mathbf{S}_{\omega,t}) - \sum_{\omega,i} \lambda_{\omega,i} (\mathbf{e}_i^T A_\omega^H A_\omega \mathbf{e}_i) \end{aligned} \tag{4}$$

と表せる. ここで, m と n は周波数及び時間フレームの bin 数, $Q_{\omega,t}$ は j 番目の対角要素が $q(\omega, T_{j,t})/\sigma_j$ である対角行列, $\boldsymbol{\lambda}_{\omega,t}$ と $\lambda_{\omega,i}$ はラグランジュの未定乗数, また $\mathbf{e}_i$ は i 番目の要素が 1 で他が 0 である基底ベクトルである.

---

*Harmonic Component Analysis for Underdetermined Blind Source Separation of Music Signals. by HOSOYA, Hiroshi (Faculty of Engineering, The University of Tokyo), IZUMI, Yosuke, ONO, Nobutaka, and SAGAYAMA, Shigeki (Graduate School of Information Science and Technology, The University of Tokyo)

4　反復推定による解法

本問題で推定するパラメータは $\mathbf{S}_{\omega,t}$, A_ω, $T_{j,t}$, σ_j^2 である．これらの直接解は解析的に求まらないため，各変数を固定して逐次に更新する反復推定を行う．各ステップで目的関数の単調増加が保証されるため，局所最適解に収束するアルゴリズムが実現できる．

以下では簡単のため，瞬時混合を想定し，A_ω を ω に依らない実数行列 A で定義して更新式を導出するが，畳み込み混合の場合も同様の更新式を導くことができる．また分散 σ_j^2 も，音源に依らない値 σ^2 とした．

4.1　音源信号 $\mathbf{S}_{\omega,t}$，分散 σ^2 の更新

(4) 式を各パラメータで偏微分することで，目的関数最大化を実現する更新式が導出できる．導出過程は省略し更新式のみ記す (ただし l は音源数)．

$$\mathbf{S}_{\omega,t} = Q_{\omega,t}^{-2} A^T (A_\omega Q_{\omega,t}^{-2} A^T)^{-1} \mathbf{O}_{\omega,t} \tag{5}$$

$$\sigma^2 = \frac{1}{mnl} \sum_{j,\omega,t} |S_{j,\omega,t}|^2 \sin^2 \frac{\omega T_{j,t}}{2} \tag{6}$$

4.2　混合行列 A の更新

目的関数 J に $\mathbf{S}_{\omega,t}$ の最尤値 (5) を代入して得られた関数を A に関する目的関数とし，これを最大化する A を求めることで A の更新を行った．解析的には陽に求まらないため，勾配法により解を求めた．目的関数の具体形は

$$J_A = -\sum_{\omega,t} \mathbf{O}_{\omega,t}^H (A Q_{\omega,t}^{-2} A^T)^{-1} \mathbf{O}_{\omega,t} - \sum_{i=1} \lambda_i (\mathbf{e}_i^T A^T A \mathbf{e}_i) - mn \sum_j \log \sigma_j^2 \tag{7}$$

と表され，J_A の A による偏微分は

$$\frac{\partial J_A}{\partial A} = -AD + 2\mathrm{Re}\{\sum_{\omega,t} (A Q_{\omega,t}^{-2} A^T)^{-1} \mathbf{O}_{\omega,t} \mathbf{O}_{\omega,t}^H (A Q_{\omega,t}^{-2} A^T)^{-1} A Q_{\omega,t}^{-2}\} \tag{8}$$

と導出できる．ただし，D は i 番目の対角成分が λ_i となる対角行列である．

4.3　ピッチ周期 $T_{j,t}$ の更新

目的関数 J の $T_{j,t}$ に関する項を $J_{T_{j,t}}$ とおくと，

$$J_{T_{j,t}} = -\sum_{\omega,t} \mathbf{S}_{\omega,t}^H Q_{\omega,t}^2 \mathbf{S}_{\omega,t} = \frac{1}{2\sigma^2} \sum_{\omega} |S_{j,\omega,t}|^2 e^{\mathrm{j}\omega T_{j,t}} + J_{else} \tag{9}$$

と展開できる (ただし，J_{else} は $T_{j,t}$ に依らない項)．$T_{j,t}$ の更新は離散値全探索によって行ったが，(9) 式からわかるように，高速フーリエ変換アルゴリズムを利用した高速な探索が可能となった．

5　シミュレーション評価実験

前章で述べたアルゴリズムの評価実験を行った．MIDI 形式のデータから作成した三つの音源信号を合成し，瞬時混合のステレオ音楽信号を作り，入力とした．サンプリング周期 16kHz，フレーム長 1024，シフト 512，窓関数を Hamming 窓とする短時間フーリエ変換により観測信号の時間周波数表現を得た．

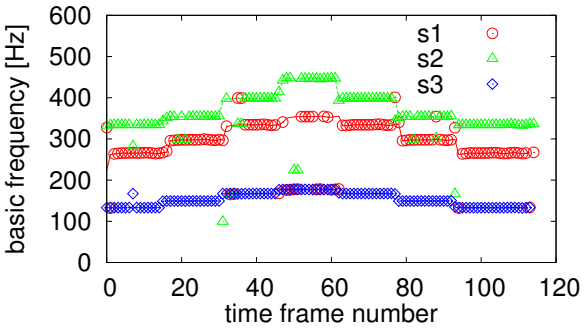

Fig. 1　ピッチ周波数の推定結果 (点は推定値，線は真値を表し，音源を色で区別)

Table 1　混合行列 A の推定結果

要素	(1,1) (2,1)	(1,2) (2,2)	(1,3) (2,3)
推定値	0.995 0.098	0.862 0.507	0.275 0.961
真値	0.985 0.174	0.866 0.500	0.342 0.940

Table 2　音源分離性能 (SN 比改善値)[dB]

音源	s1	s2	s3
SN 比改善値	7.63	2.78	3.48

混合行列の推定結果を Table 1 , 分離前後の SN 比の改善値を Table 2(観測信号の SN 比の平均値を分離前の値とした.), 推定されたピッチ周期をピッチ周波数に変換したものを Fig. 1 に示す。

Table 1, Table 2 から，提案法が音源分離手法として機能していることは確認できるが，分離性能は高くない．Fig. 1 より，ピッチ周期の推定が誤っている箇所が見られることから，アルゴリズムが局所解に陥っている可能性があり，解の探索方法の改良による改善が期待できる．

6　おわりに

本稿では，音楽信号の調波性を利用して音源分離を実現する新たな手法，調波成分分析について述べた．今後の課題としては，分離性能と計算時間を向上するために，調波性の評価指標や解の探索方法の変更が考えられる．また，非調波楽音への対応も検討したい．

参考文献

[1] A. Hyvärinen *et al.*, "Independent Component Analysis," John Wiley, New York, 2001.

[2] E. Vincent *et al.*, "First Stereo Audio Source Separation Evaluation Campaign: Data, Algorithms and Results," Proc. Int. Conf. on Independent Component Analysis and Blind Source Separation (ICA), Springer, pp. 552-559, 2007.