

調波構造テンプレートに基づく 独立ベクトル分析によるブラインド音源分離の検討*

小野拓磨 (東大院・情報理工), 亀岡弘和 (東大院・情報理工/NTT CS 研),
小野順貴 (NII), 嵯峨山茂樹 (東大院・情報理工)

1 はじめに

様々な雑音が重畳する実環境において目的とする音源信号のみを取得するための枠組みとして、マイクロホンアレーを用いた音源分離技術の研究が行われている。中でも独立ベクトル分析 (IVA: Independent Vector Analysis)[1] はパーミュテーション問題と呼ばれる分離音源の周波数間での不一致が起きにくい有効な枠組みとして知られる。独立ベクトル分析では、音源はどの周波数 bin でも共通の振幅包絡を持つと仮定されるが、一般に対象となる音源は音楽や音声など調波性という特別な性質を持つ場合がある。このとき、調波成分間の谷の部分での振幅包絡は必ずしも他の周波数 bin と一致しない一方、基本周波数や倍音間のパワーの依存関係は音源分離の重要な手がかりになり得る。そこで本研究では独立ベクトル分析に調波性を組み込んだ音源の生成モデルベースの分離手法を提案する。同時に単旋律楽器を音源とした評価実験により有用性を示す。

2 問題の定式化と調波性音源の生成モデル

時間周波数表現を用いて観測信号 $\mathbf{X}_{\tau\omega}$ に対し、復元信号を

$$\mathbf{Y}_{\tau\omega} = \mathbf{W}_{\omega}\mathbf{X}_{\tau\omega} \quad (1)$$

とモデル化し、分離行列 \mathbf{W}_{ω} の推定を考える。 $\mathbf{Y}_{\tau\omega}$ の生成モデルとして同一音源の周波数間の共起関係とパワーの非定常性を考慮でき、数式的な扱いの容易な確率モデルである時変 Gauss モデル [2]

$$Y_{m\tau\omega}|\sigma_{m\tau}^2 \sim \mathcal{N}_c(Y_{m\tau\omega}|0, \sigma_{m\tau}^2) \quad (2)$$

を仮定する。ここで $\mathcal{N}_c(\cdot|\mu, \sigma^2)$ は平均 μ 、分散 σ^2 の複素 Gauss 分布である。分散 $\sigma_{m\tau}^2$ を全周波数 bin で共通に持つことで同一音源内では類似した振幅包絡を持たせることが可能なため周波数ごとの分離とパーミュテーション整合が同時に実現できる。また、Gauss 分布の分散の時間変化を許すことで音源に Laplace 分布を仮定した IVA 同様に全フレームで見れば優 Gauss な信号モデルを仮定したことになる。

従来の IVA ではどの周波数間の振幅包絡も同程度に類似する、というモデルのため、調波成分間の谷で他音源が混入、パーミュテーション不整合が起こり得る。そこで音源モデルには、調波構造を表す櫛形のパワースペクトル密度を設定することが望ましい。そのためには基本周波数の情報が必要となるが、音源の基本周波数の情報は通常観測することができない。そこで、各フレームにおける基本周波数を潜在変数と見なし、潜在変数 (基本周波数インデックス $z_{m\tau}$) に応じて調波構造形のパワースペクトルテンプレートが一つ選択され、時変 Gauss モデルの分散が

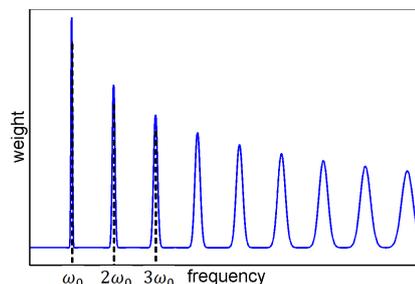


Fig. 1 調波構造テンプレート $\lambda_{\omega}^{(l)}$ の例 ($\omega = \omega_0$)

決定されるという生成プロセスを仮定する。ここで $l = 1, \dots, L$ は調波構造テンプレートのインデックスを表すこととする。以上より、 $\lambda_{\omega}^{(l)}$ を Fig. 1 のようなある基本周波数の調波構造テンプレートとすると、 $z_{m\tau}$ が与えられたもとでの音源の生成モデルは

$$Y_{m\tau\omega}|z_{m\tau}, \sigma_{m\tau}^2 \sim \mathcal{N}_c(Y_{m\tau\omega}|0, \lambda_{\omega}^{(z_{m\tau})}\sigma_{m\tau}^2) \quad (3)$$

と表される。

また、潜在変数 $z_{m\tau}$ の選ばれやすさを意味する $z_{m\tau}$ の事前確率は離散分布 $p(z_{m\tau}|\pi_{m\tau}) = \pi_{m\tau}^{(z_{m\tau})}$ とし、更にスパースな解に誘導する目的でハイパーパラメータ $\pi_{m\tau}$ のハイパー事前分布を Dirichlet 分布

$$p(\pi_{m\tau}) \propto \prod_l \{\pi_{m\tau}^{(l)}\}^{\beta-1} \quad (4)$$

と仮定する。ここで $0 < \beta \leq 1$ で、Dirichlet 事前分布により一つの $z_{m\tau}$ の事後確率のみを大きくでき単一ピッチの仮定をより強く組み込むことが可能である。

3 補助関数法による反復更新式の導出

3.1 目的関数

対数事後確率を目的関数とすれば、結局

$$\begin{aligned} & \log p(\mathbf{\Pi}, \mathbf{\Sigma}, \mathbf{W}|\mathbf{X}) \\ & \stackrel{c}{=} \sum_{m,\tau,\omega} \log \sum_{z_{m\tau}} p(Y_{m\tau\omega}|z_{m\tau}, \mathbf{\Theta})p(z_{m\tau}|\pi_{m\tau}) \\ & + \sum_{m,\tau,l} \log p(\pi_{m\tau}^{(l)}) + T \sum_{\omega} \log \det |\mathbf{W}_{\omega}| \end{aligned} \quad (5)$$

を最大化を考えればよい。ただし、 $\mathbf{X} = \{X_{m\tau\omega}\}$, $\mathbf{\Pi} = \{\pi_{m\tau}^{(l)}\}$, $\mathbf{\Sigma} = \{\sigma_{m\tau}^2\}$, $\mathbf{W} = \{\mathbf{W}_{\omega}\}$, $\mathbf{\Theta} = \{\mathbf{\Pi}, \mathbf{\Sigma}, \mathbf{W}\}$ である。

3.2 補助関数の設計

式 (5) の目的関数の大域最適解は解析的に得ることはできないが、以下に示す補助関数法により局所

* Harmonic-template-based independent vector analysis for blind source separation. by ONO Takuma (The University of Tokyo), KAMEOKA Hirokazu (The University of Tokyo/NTT CS Lab), ONO Nobutaka (NII), SAGAYAMA Shigeki (The University of Tokyo)

解を効率的に反復探索することができる。ここで、 $\sum_l \gamma_{m\tau\omega}^{(l)} = 1$ となる補助変数 $\gamma_{m\tau\omega}$ を用いれば

$$Q(\Pi, \Sigma, \mathbf{W}, \Gamma) = \sum_{m,\tau,\omega} \sum_{z_{m\tau}} \gamma_{m\tau\omega}^{(z_{m\tau})} \log \frac{p(Y_{m\tau\omega} | z_{m\tau}, \Theta) p(z_{m\tau} | \pi_{m\tau})}{\gamma_{m\tau\omega}^{(z_{m\tau})}} + \sum_{m,\tau,l} \log p(\pi_{m\tau}^{(l)}) + T \sum_{\omega} \log \det |W_{\omega}| \quad (6)$$

と補助関数を設計できる。

3.3 補助変数の更新

補助変数についての更新式は $\sum_l \gamma_{m\tau\omega}^{(l)} = 1$ の制約のもと、 $\partial Q / \partial \gamma_{m\tau\omega}^{(l)} = 0$ を解けば

$$\gamma_{m\tau\omega}^{(l)} = \frac{\pi_{m\tau}^{(l)} \mathcal{N}_c(Y_{m\tau\omega} | 0, \lambda_{\omega}^{(l)} \sigma_{m\tau}^2)}{\sum_{l'} \pi_{m\tau}^{(l')} \mathcal{N}_c(Y_{m\tau\omega} | 0, \lambda_{\omega}^{(l')} \sigma_{m\tau}^2)} \quad (7)$$

を得る。

3.4 パラメータの更新

$\partial Q / \partial \pi_{m\tau}^{(l)} = 0, \partial Q / \partial \sigma_{m\tau}^2 = 0$ より

$$\pi_{m\tau}^{(l)} = \frac{\sum_{\omega} \gamma_{m\tau\omega}^{(l)} + \beta - 1}{\sum_l \sum_{\omega} \gamma_{m\tau\omega}^{(l)} + L(\beta - 1)} \quad (8)$$

$$\sigma_{m\tau}^2 = \frac{\sum_{l\omega} \frac{\gamma_{m\tau\omega}^{(l)} |Y_{m\tau\omega}|^2}{\lambda_{\omega}^{(l)}}}{\sum_{l\omega} \gamma_{m\tau\omega}^{(l)}} \quad (9)$$

が求める更新式である。

W_{ω} は補助関数 IVA[3] 同様に 1 行ずつ更新し

$$\tilde{\sigma}_{m\tau\omega}^2 = \frac{\sigma_{m\tau}^2}{\sum_l \gamma_{m\tau\omega}^{(l)} / \lambda_{\omega}^{(l)}}, \quad (10)$$

$$V_{m\omega} = \frac{1}{T} \sum_{\tau} \left(\frac{\mathbf{X}_{\tau\omega} \mathbf{X}_{\tau\omega}^H}{\tilde{\sigma}_{m\tau\omega}^2} \right), \quad (11)$$

$$\mathbf{w}_{m\omega} = (W_{\omega} V_{m\omega})^{-1} \mathbf{e}_m, \quad (12)$$

$$\mathbf{w}_{m\omega} = \mathbf{w}_{m\omega} / \sqrt{\mathbf{w}_{m\omega}^H V_{m\omega} \mathbf{w}_{m\omega}} \quad (13)$$

とすればよい。

4 シミュレーション実験

4.1 実験目的および条件

提案手法の有用性を示す目的で、単旋律楽器を音源として用いたシミュレーションによる実験を行った。Table 1 に実験条件を示す。予め録音された残響時間の異なる二種類の室内インパルス応答 [4] を単旋律楽器音 (トランペット、サクソ) とそれぞれ畳み込み加算することで観測信号とした。部屋 E2A (残響時間 $T_{60}=300$ ms) では、 $\{-20^\circ, 0^\circ\}$ の方向から、部屋 JR2 ($T_{60}=470$ ms) では、 $\{-30^\circ, 30^\circ\}$ の方向から音源が到来するとした。

提案法に用いた重みテンプレートは 220Hz (A3) から半音ごとに 24 種類の倍音成分が $1/\sqrt{n}$ で減衰する調波構造テンプレートと全ての周波数で同じ重みのテンプレート 1 種類の合計 25 種類を用いた。提案法の分離行列の初期値として、時変 Gauss 分布に基づ

Table 1 シミュレーション実験の条件

room	E2A	JR2
reverberation time	300 ms	470 ms
frame length	1024 points (64ms)	2048 points (128ms)
number of sources	2	
number of microphones	2	
microphone spacing	28.3 mm	
impulse responses	RWCP Sound Scene Database[4]	
sampling rate	16 kHz	
data length	5 s	
frame shift	3/4 overlap	
evaluative criterion	SDR (Signal-to-Distortion)[5]	

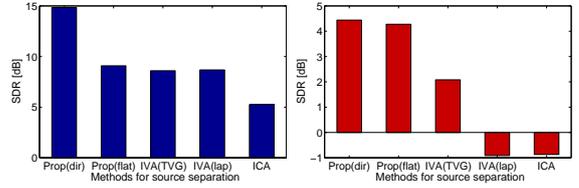


Fig. 2 SDR による部屋 E2A ($T_{60}=300$ ms, 左図) と部屋 JR2 ($T_{60}=470$ ms, 右図) における分離性能比較: 左 2 つが提案法で事前分布を Dirichlet 分布、一様事前分布とした場合

く独立ベクトル分析により分離行列を 40 回更新した値を用い、その後に提案法を 40 回反復し、復元信号を得た。提案法は $\pi_{m\tau}^{(l)}$ の事前分布として Dirichlet 分布を考慮した手法 (Prop(dir)) と事前分布を一様にした手法 (Prop(flat)) の 2 種類を行った。Dirichlet 分布を仮定した実験では、フレーム長が大きくなると事前分布の効用が小さくなるため効果を調節する目的で $p(\pi_{m\tau})$ を周波数 bin 数乗した。また、提案法 40 回反復の間に $p(\pi_{m\tau})$ のパラメータ β を 1 から 0.95 まで徐々に小さくするアニーリング処理を行った。このとき更新の間に $\pi_{m\tau}^{(l)}$ が負になることがあったが、その場合十分小さい正数に置き換えた。

比較として音源の生成モデルに時変 Gauss 分布を仮定した独立ベクトル分析による手法 (IVA(TVG)) 音源の生成モデルにラプラス分布を仮定した独立ベクトル分析による手法 (IVA(lap))[1]、独立成分分析による手法 (ICA)[6] の 3 種類をそれぞれ分離行列の更新に 80 回行った。

4.2 実験結果

SDR[5] による評価の結果を Fig. 2 に示す。どちらの環境においても従来法に比べ提案法が優位な結果を得られた。特に、部屋 E2A において Dirichlet 事前分布を考慮することで、6dB 程度の改善が見られた。

5 まとめと今後の展望

本研究では、音源の時変 Gauss モデルに調波性を示す重みテンプレートを考慮することで、単旋律楽器に対してより高精度な分離を実現した。今後の展望として、ピッチの滑らかな時間変化を隠れマルコフモデルにより表現することが挙げられる。

参考文献

- [1] T. Kim *et al.*, *Proc. ICA*, pp.165–172, 2006.
- [2] T. Nakatani *et al.*, *Proc. ICASSP*, pp. 85–88, 2008.
- [3] N. Ono, *Proc. WASPAA*, 2011.
- [4] S. Nakamura *et al.*, *Proc. LREC*, pp. 965–968, 2000.
- [5] E. Vincent *et al.*, *Trans. ASLP*, pp.1462–1469, 2006.
- [6] H. Sawada *et al.*, *Proc. ICASSP*, pp. 381–384, 2003.