HVA：調波ベクトル分析 *

○矢田部浩平 (早稲田), 北村大地 (香川高専)

1 まえがき

N 個の音源が M 本のマイクロホンによって観測されている状況を，時間周波数領域において $\mathbf{x}[t,f] \approx A[f]\mathbf{s}[t,f]$ と近似し，分離フィルタ $W[f]$ を用いて

$$W[f]\mathbf{x}[t,f] \approx W[f]A[f]\mathbf{s}[t,f] = \mathbf{s}[t,f]$$

と分離する問題を考える．この問題に対する一般的な戦略は，音源信号に対する何らかの先験情報を関数 $\mathcal{P}$ として導入し，分離フィルタに関する最適化問題

$$\underset{\{W[f]\}_{f=1}^{F}}{\text{Minimize}} \ \mathcal{P}(W[f]\mathbf{x}[t,f]) - \sum\nolimits_{f=1}^{F} \log|\det(W[f])|$$

を解くことであり，独立成分分析 (ICA) や独立ベクトル分析 (IVA) など様々な手法が提案されている．

我々はこの問題に対し，近接分離最適化を適用したアルゴリズムを提案している [1, 2]．提案法の特徴は，ペナルティ関数 $\mathcal{P}$ に関する最小化を，近接作用素と呼ばれる「より簡単な部分問題」に置き換えることで，多くの音源モデルを単一のアルゴリズムで統一的に扱える点にある．さらに，アルゴリズムの解釈を緩和することで，音源モデルを「一般の時間周波数マスキング手法」で間接的に与える拡張手法も提案している [3, 4]．**Algorithm 1** に示すように，音源を強調するマスク生成関数 $\mathcal{M}_\theta(\cdot)$ さえ定義すれば，それを代入するのみで新たな音源分離手法を実現でき，定式化や解法に囚われない提案が可能である．

本稿では，提案アルゴリズムの応用例として，信号の調波構造に基づいて分離を行う「調波ベクトル分析」を提案する [5]．ケプストラムをスパースにする閾値処理によって生成した Wiener 型マスクを新たに定義し，アルゴリズムに代入することで，音声と楽音どちらにも有効な音源分離手法を実現した．

2 時間周波数マスキングに基づく音源分離

観測データを行列 X で表現し，分離フィルタをベクトル化して $\mathbf{w}$ とすると，音源分離問題は $\mathcal{P}(X\mathbf{w})$ と $-\log|\det(\mathsf{mat}(\mathbf{w})[f])|$ の最小化なので，それらの和に対して近接分離アルゴリズムを適用し [1, 2]，$\mathcal{P}$ の近接作用素を時間周波数マスク $\mathcal{M}_\theta(\cdot)$ で置き換えることで **Algorithm 1** が得られる [3, 4]．ここで $\mathcal{M}_\theta(\cdot)$ は，音源モデルを事前分布として Gauss 雑音を除去する MAP 推定に対応しており [3, 4]，そのような雑音除去効果があれば，どんなマスクでも音源分離アルゴリズムを実現可能である．よって，新手法の開発をマスク設計問題として捉えなおすことができる．

Algorithm 1 Masking-based PDS-BSS

Input: X, $\mathbf{w}^{[1]}$, $\mathbf{y}^{[1]}$, μ_1, μ_2, α
Output: $\mathbf{w}^{[K+1]}$
for $k = 1, \ldots, K$ **do**
  $\widetilde{\mathbf{w}} = \mathrm{prox}_{\mu_1 \mathcal{I}}[\, \mathbf{w}^{[k]} - \mu_1\mu_2 X^H \mathbf{y}^{[k]} \,]$
  $\mathbf{z} = \mathbf{y}^{[k]} + X(2\widetilde{\mathbf{w}} - \mathbf{w}^{[k]})$
  $\widetilde{\mathbf{y}} = \mathbf{z} - \mathcal{M}_\theta(\mathbf{z}) \odot \mathbf{z}$
  $\mathbf{y}^{[k+1]} = \alpha\widetilde{\mathbf{y}} + (1-\alpha)\mathbf{y}^{[k]}$
  $\mathbf{w}^{[k+1]} = \alpha\widetilde{\mathbf{w}} + (1-\alpha)\mathbf{w}^{[k]}$
end for

3 調波ベクトル分析 (HVA)

提案アルゴリズムにおいて IVA は，関数 $\mathcal{P}$ として各窓をグループで扱うグループスパース誘導関数を選んだ場合に対応する．例えば，$\ell_{2,1}$ノルムを選ぶと

$$\big(\mathcal{M}_\lambda(\mathbf{z})\big)_n[t,f] = \big(1 - \lambda/(\textstyle\sum_{f=1}^{F}|z_n[t,f]|^2)^{\frac{1}{2}}\big)$$

のようなマスクが導かれ，これを代入すれば球対称 Laplace 分布に基づく IVA が実現される．

このマスクにも表れているように，IVA は信号エネルギーの時間変動に基づいて分離を行う手法であり，信号の周波数的な特徴については考慮していない．その点が，独立低ランク行列分析 (ILRMA) など信号の周波数構造を考慮できる発展的手法に分離性能で劣る原因であると考えられる．そこで本稿では，信号の調波構造を強調するマスクを新たに定義することで，調波ベクトル分析 (HVA) を提案する [5]．

3.1 ケプストラムの閾値処理

調波的な信号を解析する手段として，ケプストラムが広く知られている．調波信号の対数振幅スペクトルは周期的構造を有するので，その Fourier 変換であるケプストラム係数はスパースになる．従って，ケプストラムをスパースに誘導することで，逆変換後の振幅スペクトルの調波構造を強調することができる．

スペクトログラムの周波数軸に沿った Fourier 変換を $\mathscr{F}_f$ で表し，スパースに誘導する閾値処理 $\mathcal{T}_{\mathrm{sp}}^{\lambda}$ を組み合わせた Fourier 閾値処理作用素 $\mathcal{T}_{\mathscr{F}}^{\lambda}$ を

$$\mathcal{T}_{\mathscr{F}}^{\lambda}(\mathbf{z}) = \mathscr{F}_f^{-1}(\mathcal{T}_{\mathrm{sp}}^{\lambda}(\mathscr{F}_f(\mathbf{z})))$$

---

*HVA: Harmonic Vector Analysis. By Kohei YATABE (Waseda University) and Daichi KITAMURA (National Institute of Technology, Kagawa College).
本研究は科研費 19K20306 と関連している．

と定義する．対数振幅スペクトログラムにこの閾値処理を施すことでケプストラムをスパースに誘導でき，指数関数で対数の効果を打ち消せば，調波構造が強調された振幅スペクトログラムが手に入る．

3.2 独立でない Wiener 型マスクによる BSS

統計的独立性に基づく BSS では，音源モデル $\mathcal{P}$ は音源毎に独立な関数 $\mathcal{P} = \sum_n \mathcal{P}_n$ として定義されるので，対応するマスクも音源毎に独立に計算される．一方，音源を分離するにあたり，独立なモデルに定式化を制限する必要はないので，音源信号同士の関係性を利用した処理を考えても良いはずである．そこで，音源信号の推定値 $\hat{\mathbf{s}}$ を用いた Wiener 型マスク

$$(\mathcal{M}_{\mathrm{WL}}(\hat{\mathbf{s}}))_n[t,f] = \frac{|\hat{s}_n[t,f]|^2}{\sum_{n=1}^{N} |\hat{s}_n[t,f]|^2}$$

による，独立性の枠組みに囚われない BSS アルゴリズムを提案する．チャンネル間の相対的な関係に基づいて振幅を操作できるので，「信号を選り分ける」のに適した処理になることが期待される．

3.3 HVA のマスク関数

ケプストラムの閾値処理によって調波構造が強調された二乗振幅スペクトログラム $\upsilon^{\mathbf{z},\lambda,\varepsilon}$ を用いて

$$(\mathcal{M}^{\lambda,\varepsilon}_{\mathrm{HVA}}(\mathbf{z}))_n[t,f] = \frac{\upsilon_n^{\mathbf{z},\lambda,\varepsilon}[t,f]}{\sum_{n=1}^{N} \upsilon_n^{\mathbf{z},\lambda,\varepsilon}[t,f]}$$

のように Wiener 型マスクを考え，これを **Algorithm 1** の 6 行目に代入した BSS アルゴリズムを HVA と定義する．ただし，入力スペクトログラム $\mathbf{z}$ に対して

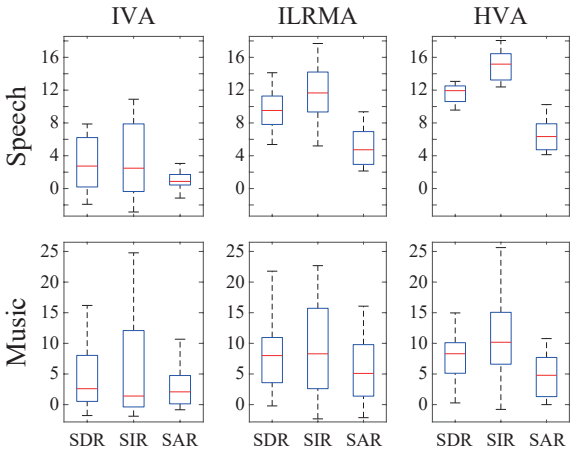
$$\rho_n^{\mathbf{z},\varepsilon}[t,f] = \log(|z_n[t,f]| + \varepsilon) - \mu_n^{\mathbf{z},\varepsilon}[t]$$

のように対数振幅を取った上でその平均値 $\mu_n^{\mathbf{z},\varepsilon}[t] = (1/F)\sum_{f=1}^{F} \log(|z_n[t,f]| + \varepsilon)$ を減算し，閾値処理をしてから平均を足し戻した対数振幅スペクトログラム

$$\varrho_n^{\mathbf{z},\lambda,\varepsilon}[t,f] = (\mathcal{T}^{\lambda}_{\mathscr{F}}(\boldsymbol{\rho}^{\mathbf{z},\varepsilon}))_n[t,f] + \mu_n^{\mathbf{z},\varepsilon}[t,f]$$

に指数関数を適用して $\upsilon_n^{\mathbf{z},\lambda,\varepsilon}[t,f] = \exp(2\varrho_n^{\mathbf{z},\lambda,\varepsilon}[t,f])$ と計算する．$\lambda > 0$ は閾値処理のパラメータを表し，$\varepsilon > 0$ は $\log(0) = -\infty$ を防ぐ小さな定数である．対数振幅スペクトルの平均値を減算することで，ケプストラムがよりスパースになるよう配慮した．

音源信号が調波構造を有する場合，HVA のマスクは調波成分に対応する周波数を強調するよう働くので，周波数間のパーミュテーションを解決できることが期待される．また，スペクトログラム全体を同時に計算に用いる ILRMA などと異なり，HVA は IVA のように各時刻で独立に処理を行うので，オンライン化に適していると考えられる．



図–1 各分離手法による SDR・SIR・SAR の改善量．

4 数値実験

提案手法の有効性を確認するために，ILRMA 提案論文 [6] と同様な実験を行った．音源として音声と楽音を用い，2 チャンネルの観測信号に対して提案法を適用し，IVA・ILRMA と性能を比較した．

音声信号として，SiSEC 2011 の劣決定分離タスクで配布されている `dev1` と `dev2` の `liverec` に含まれる 12 ファイルを観測信号として用い，最初の 2 チャンネルを選んで音源数と観測数を揃えた [6]．音楽信号は，SiSEC 2011 の楽音分離タスクで配布されている `dev1` と `dev2` に含まれる 5 曲から 2 ファイルずつ選び，RWCP データベースのインパルス応答 `E2A` と `JR2` を畳み込んだ計 10 ファイルを観測信号とした [6]．IVA と ILRMA は補助関数法で実装し，全ての手法で反復回数は 100 回に揃えた．

分離された音声 24 信号と楽音 20 信号それぞれについて SDR・SIR・SAR の改善量で評価し，ボックスプロットにまとめたものを図–1 に示す．楽音の分離において HVA は ILRMA と同等の性能を示し，音声に関しては HVA の方が高い性能であることが示唆された．また，HVA の方が ILRMA より計算時間が短く，HVA は低演算量で高性能な手法であると言える．

参考文献

[1] 矢田部浩平, 北村大地, “近接分離最適化によるブラインド音源分離,” 日本音響学会講演論文集, pp. 431–434 (2018).

[2] K. Yatabe and D. Kitamura, “Determined blind source separation via proximal splitting algorithm,” *Proc. IEEE ICASSP*, pp. 776–780 (2018).

[3] 矢田部浩平, 北村大地, “一般の時間周波数マスキングに基づく独立ベクトル分析,”日本音響学会講演論文集, pp. 219–220 (2018).

[4] K. Yatabe and D. Kitamura, “Time-frequency-masking-based determined BSS with application to sparse IVA,” *Proc. IEEE ICASSP*, pp. 715–719 (2019).

[5] K. Yatabe and D. Kitamura, “Determined BSS based on time-frequency masking and its application to harmonic vector analysis,” — (submitting).

[6] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization,” *IEEE/ACM Trans. ASLP*, **24**(9), 1626–1641 (2016).