

## 補助関数法に基づく幾何学的制約付き独立ベクトル分析の 車室内音声強調への適用\*

☆後藤加奈, 高橋理希, 李莉, 牧野昭二, 山田武志 (筑波大)

### 1 はじめに

離れた位置にあるマイクロフォンを用いて音声をキャプチャする場合, 雑音や干渉音により音声の品質が著しく低下するため, 雑音の多い環境での音声強調技術の必要性が高まっている. 例えば高速で移動する車室内では, エンジンや風などの雑音によって搭乗者間のコミュニケーションの難易度が上がり, ナビゲーションシステムの音声認識精度が低下する. このような車室内コミュニケーションの状況を改善するために, 目的音声に重畳した雑音を取り除くことを目的とした音声強調手法が広く研究されている. 車室内のマイクロフォンに対する話者の方向が一定であることを考えると, 一定方向からの信号を強調したり抑制したりするビームフォーミング [1, 2] を用いることは合理的である. しかしこの方法では, 高い強調性能を達成するために多数のマイクロフォンが必要となり, 音声と雑音の両方を学習しなければならない.

その他の指向性を形成する音声強調手法には, 幾何学的制約付きブラインド音源分離 (Blind Source Separation; BSS) [3, 4] があり, 信号の空間情報やセンサ位置といった事前情報を幾何学的制約として与え分離行列を求めることで, 混合信号から目的方向の成分のみを分離して得る手法である. 幾何学的制約付き BSS は, 音源信号の統計的独立性に基づいて空間的に推定されたヌルを用いて信号を分離するため, 訓練サンプルを必要とせず, 限られた数のマイクロフォンで高い性能を発揮する. 幾何学的制約付き独立ベクトル分析 (Geometrically Constrained Independent Vector Analysis; GCIVA) [5, 6] はそのような手法の一つであり, 幾何学的制約と独立ベクトル分析 (IVA) の最適化問題を組み合わせた手法である. [6] では, 補助関数法とベクトルワイズ座標降下法 (VCD) に基づいて GCAV-IVA と呼ばれる幾何学的制約付き BSS の手法が導出されており, 計算コストが低いため収束が速いことや, ステップサイズの調整が不要であることが注目されている.

[6] では GCAV-IVA のシミュレーション状況下での有効性が検証されているが, 本研究では車室内の環境で実験を行い有効性を検証する [7]. 特に三角形に

配置されたマイクロフォンアレイ (三角マイク) を用いた GCAV-IVA による車室内音声強調システムを検討し, 複数の条件で測定した車室内インパルス応答を用いてテストデータを作成する. 事前情報として音源の到来方向を与えられた GCAV-IVA と, 学習データから空間情報を学習する GCAV-IVA の 2 パターンに対して性能評価を行う.

### 2 幾何学的制約付き独立ベクトル分析

#### 2.1 定式化

$I$  個のマイクロフォンで  $J$  個の音源を観測する場合を考える.  $x_i(\omega, t)$  と  $y_j(\omega, t)$  をそれぞれ  $i$  番目のマイクロフォンで観測された信号と  $j$  番目の推定音源の短時間フーリエ変換 (Short-Time Fourier Transform; STFT) 係数とする. ここで,  $\omega$  と  $t$  はそれぞれ周波数および時間フレームである. 観測信号と推定された音源の周波数ごとのベクトル表現はそれぞれ以下の式で与えられる.

$$\mathbf{x}(\omega, t) = [x_1(\omega, t), \dots, x_I(\omega, t)]^T \in \mathbb{C}^I, \quad (1)$$

$$\mathbf{y}(\omega, t) = [y_1(\omega, t), \dots, y_J(\omega, t)]^T \in \mathbb{C}^J, \quad (2)$$

ここで  $J = I$  で,  $(\cdot)^T$  は転置を表す. STFT 窓の長さが音源とマイクロフォン間のインパルス応答よりも十分に長い場合, 観測信号と推定音源の関係は時間不変瞬時混合モデルを用いて次のように表すことができる.

$$\mathbf{y}(\omega, t) = \mathbf{W}(\omega)\mathbf{x}(\omega, t), \quad (3)$$

ここで,  $\mathbf{W}(\omega) = [\mathbf{w}_1(\omega), \dots, \mathbf{w}_I(\omega)]^H$  は,  $I \times I$  の分離行列,  $(\cdot)^H$  はエルミート転置を表す.

分離行列  $\mathcal{W} = \{\mathbf{W}(\omega)\}_\omega$  は, 次の目的関数を最小化することにより推定される.

$$J_{\text{IVA}}(\mathcal{W}) = \sum_{j=1}^J \mathbb{E}[G(\mathbf{y}_j(t))] - \sum_{\omega=1}^{\Omega} \log |\det \mathbf{W}(\omega)|, \quad (4)$$

ここで  $\Omega$  は周波数ビンの数,  $\mathbb{E}[\cdot]$  は期待値演算子を表す.  $\mathbf{y}_j(t)$  は次式で定義された音源単位でのベクトル表現である.

$$\mathbf{y}_j(t) = [y_j(1, t), \dots, y_j(\Omega, t)]^T \in \mathbb{C}^\Omega. \quad (5)$$

\*Study on geometrically constrained IVA with auxiliary function approach and VCD for in-car communication. by Kana GOTO, Riki TAKAHASHI, Li LI, Shoji MAKINO, Takeshi YAMADA (University of Tsukuba)

$G(\mathbf{y}_j(t))$  は  $G(\mathbf{y}_j(t)) = -\log p(\mathbf{y}_j(t))$  の関係を持つコントラスト関数であり、 $p(\mathbf{y}_j(t))$  は  $j$  番目の音源の多変量確率密度関数である。一般的に、コントラスト関数は以下の式のように球形多変量分布が用いられる。

$$G(\mathbf{y}_j(t)) = G_R(r_j(t)), \quad (6)$$

$$r_j(t) = \|\mathbf{y}_j(t)\|_2 = \sqrt{\sum_{\omega} |y_j(\omega, t)|^2}. \quad (7)$$

さらに、IVA で推定された  $j$  番目の分離フィルタの遠距離場応答を  $\theta$  方向に制限する幾何学的制約を考えると、次のような目的関数が得られる。

$$J_c(\mathcal{W}) = \sum_{j=1}^J \lambda_j \sum_{\omega=1}^{\Omega} |\mathbf{w}_j^H(\omega) \mathbf{d}_j(\omega, \theta) - c_j|^2. \quad (8)$$

ここで、 $\mathbf{d}_j(\omega, \theta)$  は  $\theta$  方向を指すステアリングベクトル、 $c_j$  は非負の値を持つ制約、 $\lambda_j \geq 0$  は制約を重み付けするパラメータであり、これは線形拘束付最小分散型 (Linear Constraint Minimum Variance; LCMV) ビームフォーマ [9] で使用された概念である。  $c_j = 1$  の場合、空間フィルタは目的信号を保存するために従来の遅延和法 (Delay and Sum; DS) ビームフォーマのステアリングベクトルを  $\theta$  方向に強制的に形成するのに対し、 $c_j$  の値が小さい場合は、目的音源を抑制し他の全ての音源を保存することを目的とした、目的方向  $\theta$  への空間的なヌルを形成する。GCIVA の目的関数は以下のようにまとめられる。

$$J(\mathcal{W}) = J_{\text{IVA}}(\mathcal{W}) + J_c(\mathcal{W}). \quad (9)$$

## 2.2 補助関数法に基づく推論アルゴリズム

GCAV-IVA の推論アルゴリズムは、補助関数法 [10] に基づき、補助関数  $J^+(\mathcal{W}, \mathcal{V})$  が、 $J(\mathcal{W}) = \min_{\mathcal{V}} J^+(\mathcal{W}, \mathcal{V})$  を満たすように設計されている。解析的に解くのが難しい目的関数を直接最適化するのではなく、補助関数  $J^+(\mathcal{W}, \mathcal{V})$  を、 $\mathcal{W}$  と  $\mathcal{V}$  を交互に最小化することによって最適化する。

幾何学的制約は線形であるため、元の AuxIVA の補助関数 [11] と線形制約を組み合わせることで上界となる補助関数を得られる。

$$J^+(\mathcal{W}, \mathcal{V}) \triangleq \sum_{j=1}^J \sum_{\omega=1}^{\Omega} \left\{ \frac{1}{2} \sum_j \mathbf{w}_j^H(\omega) \mathbf{V}_j(\omega) \mathbf{w}_j(\omega) - \log |\det \mathbf{W}(\omega)| \right\} + J_c(\mathcal{W}), \quad (10)$$

ここで  $\mathbf{V}_j(\omega)$  は、次のように表される加重共分散である。

$$\mathbf{V}_j(\omega) = \mathbb{E} \left[ \frac{G'_R(r_j(t))}{r_j(t)} \mathbf{x}(\omega) \mathbf{x}^H(\omega) \right] \quad (11)$$

また、 $=^c$  は定数項までの等質性を表す。音源モデル  $G_R(r_j(t)) = r_j(t)$  を使用する場合、 $\mathbf{V}_j(\omega)$  は、 $\mathbb{E}[\mathbf{x}\mathbf{x}^H/r_j(t)]$  と表現することができる。

表記を簡単にするため、指標  $\omega$  と  $\theta$  を省略し導出された更新規則をまとめると、以下のようになる。

$$\mathbf{u}_j = \mathbf{D}_j^{-1} \mathbf{W}^{-1} \mathbf{e}_j, \quad (12)$$

$$\hat{\mathbf{u}}_j = \lambda_j c_j \mathbf{D}_j^{-1} \mathbf{d}_j, \quad (13)$$

$$\mathbf{h}_j = \mathbf{u}_j^H \mathbf{D}_j \mathbf{u}_j, \quad (14)$$

$$\hat{h}_j = \mathbf{u}_j^H \mathbf{D}_j \hat{\mathbf{u}}_j, \quad (15)$$

$$\mathbf{w}_j = \begin{cases} \frac{1}{\sqrt{\hat{h}_j}} \mathbf{u}_j + \hat{\mathbf{u}}_j & (\text{if } \hat{h}_j = 0), \\ \frac{\hat{h}_j}{2\hat{h}_j} \left[ -1 + \sqrt{1 + \frac{4\hat{h}_j}{|\hat{h}_j|^2}} \right] \mathbf{u}_j + \hat{\mathbf{u}}_j & (\text{o.w.}). \end{cases} \quad (16)$$

ここで、 $\mathbf{D}_j = \mathbf{V}_j + \lambda_j \mathbf{d}_j \mathbf{d}_j^H$ 、 $\mathbf{e}_j$  は単位行列  $I \times I$  の  $j$  番目の列である。詳細は [6] と [12] に記載されている。

## 3 評価実験

マイクロフォンの数と音源数がともに 3 である場合の車室内コミュニケーションにおける GCAV-IVA の有用性を評価するため、実測のインパルス応答を使用した音声強調実験を行った。

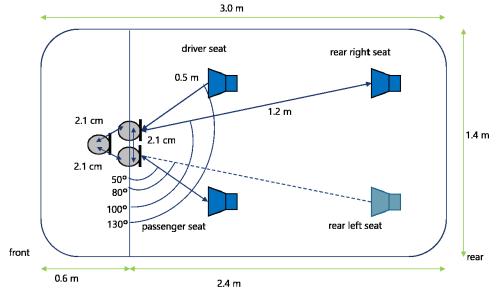
### 3.1 固定制約とデータ適応制約の GCAV-IVA

本実験では GCAV-IVA で用いるステアリングベクトルを推定するために、2 種類の方法を採用する。第 1 の方法は、マイクロフォンアレイの形状に関する事前知識を持った平面波伝播仮定の下での到来方向 (Direction of Arrival; DOA) に基づくものであり、これを **GCAV-IVA(1)** (固定制約) と呼ぶ。

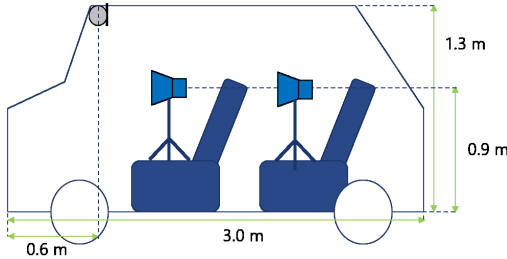
第 2 の方法は、観測された音声の共分散行列  $\mathbf{R}_s$  の固有値分解に基づくものである。推定されたステアリングベクトルは、共分散行列の主固有ベクトルとして  $\mathbf{d}(\omega) = \text{PE}\{\mathbf{R}_s\}$  と与えられる。ここで PE は、行列の主固有ベクトルを取り出す操作である。第 2 の方法は目的音声のみを含む訓練データを用いてステアリングベクトルを推定するため、**GCAV-IVA(2)** (データ適応制約) と呼ぶ。なお、適応ビームフォーマは目的音声と妨害音声の両方を学習するが、GCAV-IVA(2) は目的音声のみを学習する。

### 3.2 実験条件

本実験では、ATR デジタル音声データベースのセット B に収録されている、全 503 文の音素バランス文の男性 3 話者、女性 3 話者分の、約 20 秒間の音声ファイルのデータを使用した。車室内のマッピングに取り付けた三角マイクを用いて、運転席、助手



(a) Plan view



(b) Side view

Fig. 1: Layout of sound sources and microphones

Table 1: Experimental conditions

|                    |  |
|--------------------|--|
| マイクロフォンの数          | 3  |
| 音源数                | 3  |
| 残響時間 $T_{60}$      | 58 ms                                      |
| 目的音源の DOA          | $130^\circ$                                |
| 干渉音源の DOA          | $50^\circ, 80^\circ / 50^\circ, 100^\circ$ |
| STFT フレーム長         | 1024 samples                               |
| STFT フレームシフト幅      | 256 samples                                |
| 学習区間長<br>(目的/干渉音源) | 5 sec/5 sec                                |
| テスト区間長             | 5 sec                                      |

席，左右後部座席に配置したスピーカーで再生される TSP(Time Stretched Pulse) 信号を記録し，車室内のインパルス応答を測定した．スピーカーは各座席中央，左寄り，右寄りに置いた．インパルス応答は窓を開けている状態と閉めている状態の 2 通りで測定した．車室内とマイクロフォンアレイ配置の詳細を図 1 に示す．

測定されたインパルス応答をランダムに選択された音声に畳み込み，3 人の話者の混合信号を生成した．目的話者は運転手とし，干渉話者のうち 1 人は助手席，もう 1 人は後部座席の左右どちらかに座っているものとした．その他の実験条件を表 1 に示す．

### 3.3 手法と評価尺度

固定制約とデータ適応制約の GCAV-IVA を，AuxIVA，DS ビームフォーマ，MVDR ビームフォーマ，maxSNR ビームフォーマと比較した．これらの手法は必要とする事前情報の種類によって 3 グループに分類される．BSS グループ，目的音声の DOA を必要とする固定制約グループ，目的音（と妨害音）の

Table 2: Average SDR, SIR, and SAR [dB] over test dataset achieved by different method.

| 手法      |             | SDR          | SIR          | SAR          |
|---------|-------------|--------------|--------------|--------------|
| BSS     | AuxIVA      | 16.32        | 22.76        | 17.66        |
| 固定制約    | DS          | 0.14         | 0.15         | <b>31.13</b> |
|         | GCAV-IVA(1) | 8.06         | 12.78        | 11.30        |
| データ適応制約 | MVDR        | 17.22        | <b>29.13</b> | 17.61        |
|         | maxSNR      | 19.09        | 27.33        | 20.13        |
|         | GCAV-IVA(2) | <b>19.12</b> | 26.16        | 20.25        |

Table 3: Average SDR [dB] under conditions where car window was open or closed

| 手法      |             | 窓開          | 窓閉           |
|---------|-------------|-------------|--------------|
| BSS     | AuxIVA      | 15.01       | <b>17.63</b> |
| 固定制約    | DS          | <b>0.15</b> | 0.14         |
|         | GCAV-IVA(1) | <b>8.08</b> | 8.05         |
| データ適応制約 | MVDR        | 14.05       | <b>20.40</b> |
|         | maxSNR      | 17.37       | <b>20.82</b> |
|         | GCAV-IVA(2) | 17.31       | <b>20.94</b> |

Table 4: Average SDR [dB] under conditions where the interferer was at rear left of right seat.

| 手法      |             | 後部左          | 後部右   |
|---------|-------------|--------------|-------|
| BSS     | AuxIVA      | <b>16.65</b> | 16.00 |
| 固定制約    | DS          | <b>0.21</b>  | 0.08  |
|         | GCAV-IVA(1) | <b>8.33</b>  | 7.80  |
| データ適応制約 | MVDR        | <b>18.35</b> | 16.09 |
|         | maxSNR      | <b>19.51</b> | 18.68 |
|         | GCAV-IVA(2) | <b>19.7</b>  | 18.54 |

訓練データを必要とするデータ適応制約グループである．AuxIVA と GCAV-IVA は，単位行列で初期化された分離行列を推定するために 50 回の反復を行った．AuxIVA は BSS 方式であるため，出力チャンネルの順番は任意である．そのため全チャンネルの出力を評価し，最も高い数値を採用した．

音声強調性能の評価には，signal-to-distortion ratio (SDR)，source-to-interferences ratio (SIR)，source-to-artifacts ratio (SAR) を用いる．これらは値が大きいほど高い分離性能を示す．

### 3.4 結果と考察

表 2 に全データセットの平均 SDR，SIR，SAR スコアを示す．GCAV-IVA は各カテゴリにおいてベースラインとなるビームフォーミング手法よりも優れており，データ適応制約型 GCAV-IVA(2) は全手法の中で最高の SDR スコアを達成した．また，固定制約グループの場合，GCAV-IVA(1) は SDR と SIR の両方で大幅な改善を達成した．データ適応制約グループでは図 2 のビームパターンも類似しており，特に maxSNR と GCAV-IVA(2) の性能差は僅少であったが，事前情報として MVDR，maxSNR ビームフォーマが目的音源と干渉音源の両方の区間を必要とするのに対し，GCAV-IVA(2) は目的音源のみを必要とする．これらの結果から，GCAV-IVA が車室内音声強調システムとして有効であることが確認された．

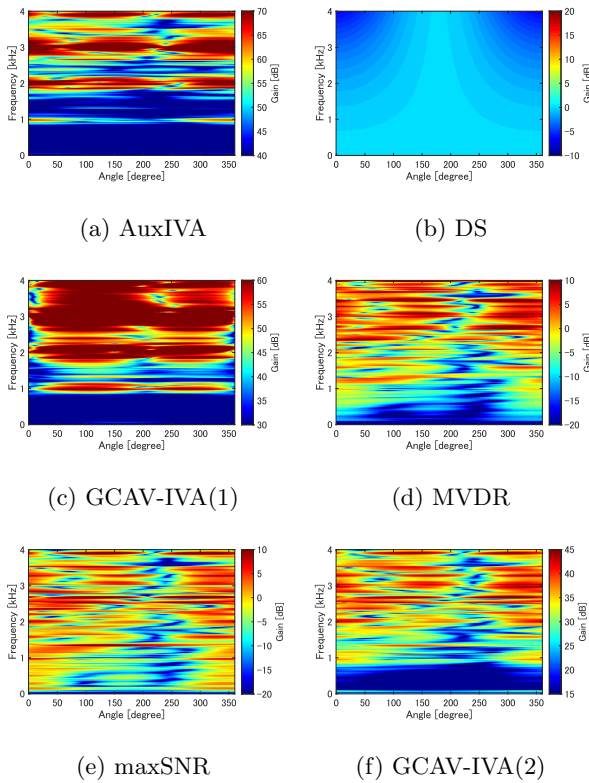


Fig. 2: Example of beam patterns obtained by each method

次に、より具体的な条件での結果を示す。表 3 に、車窓を開けている場合と閉めている場合の結果を示す。これら 2 通りの条件において、データ適応制約グループの手法で明確な差が見られ、いずれも窓を開けている状態の方が高い性能を示した。これは、窓を開けた状態ではガラスからの反響が減少するためであると考えられる。また、表 4 は、2 人目の干渉話者が後部座席の左右どちらに座っているかで結果を比較している。全ての手法において、左後席に干渉話者がいる場合の方が高い性能を示した。これは、左後席の方が目的話者方向との角度が離れていることから、空間フィルタリングを行うのにより適した条件であるためと考えられる。

#### 4 まとめ

本研究では、GCAV-IVA の車室内音声強調システムとしての有効性を検討した。GCAV-IVA は、IVA とビームフォーミングに基づく線形制約を組み合わせた指向性音声強調方式である。GCAV-IVA の制約を用いるために必要なステアリングベクトルを得るため、2 通りのアプローチを適用した。三角形マイクロフォンアレイを用いて測定した車室内インパルス応答を用いて、GCAV-IVA の音声強調実験を実施し、従来手法であるビームフォーミングや AuxIVA と比

較した。実験の結果、GCAV-IVA は全ての従来手法より高い性能を示した。

謝辞 本研究は科研費 19H04131, 戦略的基盤技術高度化支援事業の助成を受けた。

#### 参考文献

- [1] H. L. Van Trees, “Optimum array processing: Part IV of detection, estimation, and modulation theory,” 2004, John Wiley & Sons.
- [2] M. Wölfel, and J. McDonough, “Distant speech recognition,” 2009, John Wiley & Sons.
- [3] L. C. Parra, and C. V. Alvino, “Geometric source separation: Merging convolutive source separation with geometric beamforming,” *IEEE Trans. SAP*, vol. 10, no. 6, pp. 352–362, 2002.
- [4] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, “Blind source separation based on a fast-convergence algorithm combining ICA and beamforming,” *IEEE Trans. ASLP*, vol. 14, no. 2, pp. 666–678, 2006.
- [5] A. H. Khan, M. Taseska, and E. A. P. Habets, “A geometrically constrained independent vector analysis algorithm for online source extraction,” in *Proc. LVA/ICA*, pp. 396–403, 2015.
- [6] L. Li, and K. Koishida, “Geometrically constrained independent vector analysis for directional speech enhancement,” in *Proc. ICASSP*, pp. 846–850, 2020.
- [7] K. Goto, L. Li, R. Takahashi, S. Makino, and T. Yamada. “Study on Geometrically Constrained IVA with Auxiliary Function Approach and VCD for In-Car Communication,” in *Proc. AP-SIPA*, 2020.
- [8] N. Makishima, Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, “Independent deeply learned matrix analysis with automatic selection of stable microphone-wise update and fast sourcewise update of demixing matrix,” *Signal Processing (Elsevier)*, vol.178, no.107753, 12 pages, 2021.
- [9] J. Bourgeois and W. Minker, Eds., “Linearly constrained minimum variance beamforming,” pp. 27–38, 2009.
- [10] D. R Hunter and K. Lange, “A tutorial on MM algorithms,” *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.
- [11] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” in *Proc. WASPAA*, pp. 189–192, 2011.
- [12] Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, “Vectorwise coordinate descent algorithm for spatially regularized independent low-rank matrix analysis,” in *Proc. ICASSP*, pp. 746–750, 2018.