2 値マスクの併用による独立ベクトル分析のリアルタイム化

REAL-TIME COMPUTATION OF INDEPENDENT VECTOR ANALYSIS USING BINARY MASKING

太刀岡勇気　成田知宏　石井純
Yuuki Tachioka　Tomohiro Narita　Jun Ishii

三菱電機株式会社 情報技術総合研究所
Information Technology R&D Center, Mitsubishi Electric Corporation

1　まえがき

複数話者による発話の同時認識の前段には，音源分離処理が必要である．著者らは，音源分離に，音源到来方向などの物理的な情報に基づく方法と統計的な独立性を元に分離を行うブラインド音源分離法を併用することで，頑健性を向上させた [1]．ただし後者は分離行列を発話ごとに推定するため，発話終了を待たなければならないという問題があった．文献 [2] のように，目的関数を変更する以外にも，簡便にいくつかのフレームを束ねたブロックごとに分離行列を推定することでリアルタイム化が実現できるが，精度が低下する．この問題を物理的な方法の併用により回避することが，本報の目的である．

2　時間・周波数 2 値マスク (binM)

時間フレーム t，周波数ビン ω の時間差 $\tau(\omega,t)$ は x_2/x_1 の位相部分で表される．x_1，x_2 はマイク 1, 2 の観測信号の短時間フーリエ変換であり，まとめて $\boldsymbol{x}(\omega,t) = (x_1(\omega,t), x_2(\omega,t))^\top$ で表す ($\top$ は転置)．到来方向 θ に対して，マスク $W(\omega,t) = (\boldsymbol{w}_1(\omega,t), \boldsymbol{w}_2(\omega,t))^h$ は

$$\boldsymbol{w}_k(\omega,t) = \begin{cases} \epsilon \boldsymbol{e}_k & \text{if} |c/l_m \sin^{-1}\tau(\omega,t) - \theta| > \theta_c, \\ \boldsymbol{e}_k & \text{if} |c/l_m \sin^{-1}\tau(\omega,t) - \theta| \le \theta_c, \end{cases}$$

のように設定される (h は Hermite 転置)．k はマイクの ID，$\boldsymbol{e}_k$ は単位ベクトル (k 番目の要素が 1)，ϵ は小さい定数，θ_c は許容誤差，c は音速，l_m はマイク間隔である．$\boldsymbol{y}(\omega,t) = W(\omega,t)\boldsymbol{x}(\omega,t)$ より分離信号 $\boldsymbol{y}$ を得る．

3　リアルタイム独立ベクトル分析 (RT-IVA)

音源の独立性に基づく手法は，一般的に周波数ビンごとに音声を分離するため (例えば独立成分分析)，分離話者間の混同が起こる．IVA では，周波数ビンをまたがる目的関数 (1) を最小化することでこの問題を回避している．リアルタイム化のためには，いくつかのフレームを束ねたブロック b に対して，ブロック内で時不変の分離行列 $W(\omega,b)$ の組 $\boldsymbol{W}^b$ を決定する．($r^b_{k,t}$ は (2) の L_2 ノルム，E は時間に関する期待値)

$$J(\boldsymbol{W}^b) = \sum_k E[r^b_{k,t}] - \sum_\omega \log|\det W(\omega,b)|. \quad (1)$$

$\boldsymbol{W}$ の最適化のために，補助関数で J の上限を抑える方法が提案されている [3]．補助変数の更新 ((2)) 後に分離行列の更新 ((3)) を行う手順を繰り返す．ただし毎ブロック新たに $\boldsymbol{W}^b$ を推定したのでは，ブロック間でパーミュテーションが発生するため，$\boldsymbol{W}^b$ の初期値を前ブロックの推定結果 $\boldsymbol{W}^{b-1}$ とし，反復した．

$$\begin{aligned} r^b_{k,t} &= L_2\left[\boldsymbol{w}^h_k(\omega,b)\boldsymbol{x}(\omega,t)\right], \\ V_k(\omega,b) &= E\left[\boldsymbol{x}(\omega,t)\boldsymbol{x}^h(\omega,t)/r^b_{k,t}\right]. \end{aligned} \quad (2)$$

$$\begin{aligned} \boldsymbol{w}_k(\omega,b) &\leftarrow (W(\omega,b)V_k(\omega,b))^{-1}\boldsymbol{e}_k, \\ \boldsymbol{w}_k(\omega,b) &\leftarrow \boldsymbol{w}_k(\omega,b)/\sqrt{\boldsymbol{w}^h_k(\omega,b)V_k(\omega,b)\boldsymbol{w}_k(\omega,b)}. \end{aligned} \quad (3)$$

4　音声認識実験

RWCP 中の室 E2A を対象とし，線アレイの内 2 マイクを使った．l_m は表 1 の 3 通りとした．実験条件の詳細と binM と IVA の統合法は，文献 [1] を参照されたい．窓長は 60ms，フレームシフトは 30ms とした．RT-IVA のブロックサイズは 10(300ms) および 20(600ms) とした．

表 1 に各種アルゴリズムと単語正解精度の関係を示す．binM は，マイク間隔が大きくなるとエリアシングの影響で性能が低下する．IVA の性能は，マイク間隔にそれほど依存しないが，マイク間隔が狭い場合に binM に劣る．RT-IVA は，ブロックサイズを長くとると性能が向上するものの，発話全体を使ったものよりは大きく性能が低下する．binM と統合することで，RT-IVA の性能は大きく向上し，マイク間隔が 2.85 および 5.7cm の場合には発話全体を用いたものと同等となった．

5　まとめ

独立ベクトル分析のリアルタイム化を検討し，2 値マスクとの併用により音源分離の頑健性が向上した．

表 1　2 値マスク (binM)，独立ベクトル分析 (IVA)，リアルタイム IVA(RT-IVA) の単語正解精度 [%].

algorithm \ l_m	2.85[cm]	5.7[cm]	37.5[cm]
binM	84.80	76.40	36.92
IVA(batch) [3]	74.05	73.48	67.98
binM+IVA(batch) [1]	84.35	78.87	65.33
RT-IVA(10)	41.87	37.80	31.18
RT-IVA(20)	58.25	57.70	51.82
binM+RT-IVA(10)	83.93	76.88	49.87
binM+RT-IVA(20)	84.17	78.73	56.55

参考文献

[1] Y. Tachioka, T. Narita, and J. Ishii, “Semi-blind source separation using binary masking and indepedent vector analysis,” IEEJ Transactions on Electrical and Electronic Engineering, 2015. 1

[2] T. Taniguchi, N. Ono, A. Kawamura, and S. Sagayama, “An auxiliary-function approach to online independent vector analysis for real-time blind source separation,” in Proc. HSCMA, 1569905811, 2014.

[3] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” in Proc. WASPAA, 189-192, 2011.