

低遅延でオンライン動作する残響除去と音源分離の同時最適化*

☆上田哲也 (筑波大) 中谷智広 池下林太郎 木下慶介 荒木章子 (NTT) 牧野昭二 (筑波大)

1 はじめに

近年、ブラインド音源分離 (Blind Source Separation: BSS) を用いたアプリケーションが普及しつつある [1, 2]. BSS とはマイクロホンで観測された信号のみから個々の話者を表す音源信号に分離する技術であり、補聴器の改善や車室内コミュニケーションの補助など様々な応用が期待される。車室内では各座席が同じ方向を向いており、エンジン音や同乗者の発話といった複数の妨害音が存在することから、特に運転席と後部座席間での会話が成り立ちにくくなる [2]. そこで、車室内に複数のマイクロホンを設置し、マイクロホンで観測された複数の発話や雑音を含む音から BSS を用いて例えば運転手の発話だけを強調して離れた同乗者に提示して会話を補助するシステムの実用化が検討されている [3]. しかし、音源分離に遅延が生じると発話者の音声と強調した音声が発話者のように聞こえ、明瞭度が低下することが知られている [4]. その為、BSS を以上のシステムに適用するには低遅延化が要求される。また、バッチ処理のようにマイクロホンで観測された信号全体を分離するのではなく、時間的に逐次に観測される信号のみで分離するオンライン化も重要である。

マイクロホンの数が音源数以上である優決定条件下の BSS においては、音源信号間の独立性を最大化するように分離フィルタを推定する独立成分分析 (Independent Component Analysis: ICA) [5] が有用であることが知られている。周波数領域 BSS において、独立ベクトル分析 (Independent Vector Analysis: IVA) [6] は時間変化する各音源信号の分散が周波数不変であると仮定することで、音源分離問題と順序問題を同時に解決している。さらに近年では、補助関数法を用いて高速化した IVA [7, 8] やそのオンライン手法 [9] が提案されている。

周波数領域 BSS では、観測信号が入力されてからその信号が短時間フーリエ変換 (Short-Time Fourier Transform: STFT) されるまでのアルゴリズム遅延が不可避である [10]. 音源分離の補聴器や車室内コミュニケーションシステムへの適用には低遅延化が要求されるため [4, 11], STFT の窓長を短くする必要がある。しかし、IVA などの手法では、残響による分離性能の劣化を回避する為に STFT の窓長が十分に長い必要がある [12]. 一般に車室内の残響時間は短い、STFT の窓長を短くした場合、残響の影響は無視できなくなる。この問題に対して、重み付き予測誤差法 (Weighted Prediction Error: WPE) [13] を用いることで分離性能を妨げる残響を取り除くことができる。単純な方法としては、オンライン IVA [9] とオンライン WPE [14, 15] を連結することで、分離性能を維持しつつ窓長を短くすることが期待できる。しかし、この連結した手法は音源分離と残響除去で各々最適化しているため、両方を考慮した同時最適化が保証されていない。バッチ処理による同時最適化手法は提案されている [16, 17] が、残響除去の推定を行う度に巨大な逆行列演算を行う為、計算コストが大きくなる。また、これらのバッチ手法をオンライン化する方法は

未だ提案されていない。

近年、バッチ処理での同時最適化における計算コストの問題を解決する手法が提案された [18, 19]. この手法は WPE で推定する残響除去フィルタを音源の数だけ増やして各々の音源の残響を取り除くように残響除去フィルタを求める。残響除去フィルタの数が増える代わりに、1つ1つの残響除去フィルタの更新の為に計算する逆行列が従来の WPE と同じ大きさになり、総合的に見ると従来より少ない計算コストで同時最適化を実現できる。本稿はこの同時最適化手法をオンライン化した手法を提案する。提案手法は従来のオンライン手法 [9] に簡単な改良を加えるだけで実装でき、実時間動作を十分に実現できる利点がある。雑音を含めた車室内環境下での提案手法の有効性を評価する為に、アルゴリズムの詳細は [20] で述べ、本稿では音源分離実験の結果をより詳細に示す。

2 問題の定式化

M 個のマイクロホンで N 個の音源から到来する信号を観測する場合を考える。 m 番目のマイクロホンで観測される信号と n 番目の音源信号の時間周波数成分をそれぞれ $\mathbf{x}(f, t) = [x_1(f, t), \dots, x_M(f, t)]^T \in \mathbb{C}^M$ と $\mathbf{s}(f, t) = [s_1(f, t), \dots, s_N(f, t)]^T \in \mathbb{C}^N$ とする。ただし、 f と t は周波数ビンと時刻のインデックス、 F と T はその総数である。 $(\cdot)^T$ は行列の転置である。以降、 $M = N$ の場合のみを扱う。本稿では、観測信号 $\mathbf{x}(f, t)$ のみを入力として音源信号 $\mathbf{s}(f, t)$ を低遅延 (=短い窓長) かつオンライン処理で推定することをねらう。

3 ベースライン手法

本節では提案手法と比較するベースライン手法について簡潔に述べる。

3.1 オンライン IVA

オンライン IVA [9] は窓長が残響時間より十分に長いと仮定し、反復射影法 (Iterative Projection: IP [8]) を用いて音源信号 $\mathbf{s}(f, t)$ を生成する分離フィルタを推定する。しかし、この手法は低遅延化の為に STFT の窓長を短くすると残響によって仮定が成り立たなくなり、分離性能が劣化してしまう。

3.2 WPE+IVA (separate)

観測信号から残響を取り除く手法として、WPE が提案されている [13–16]. この手法は各周波数ビンで予測誤差を最小にするように多チャンネル線形フィルタを求めて観測信号に含まれる残響を取り除く。

オンライン WPE [15] とオンライン IVA [9] を結合したフローを図 1(a) に示す。オンライン WPE の出力をオンライン IVA に入力することで分離性能を妨げる残響成分を取り除きつつ短い窓長でより高い分離性能が期待できる。以降この連結した手法を WPE+IVA (separate) と呼ぶ。

*Joint optimization of source separation and dereverberation that operates with online and low latency. by Tetsuya UEDA (University of Tsukuba), Tomohiro NAKATANI, Rintaro IKESHITA, Keisuke KINOSHITA, Shoko ARAKI (NTT), Shoji MAKINO (University of Tsukuba)

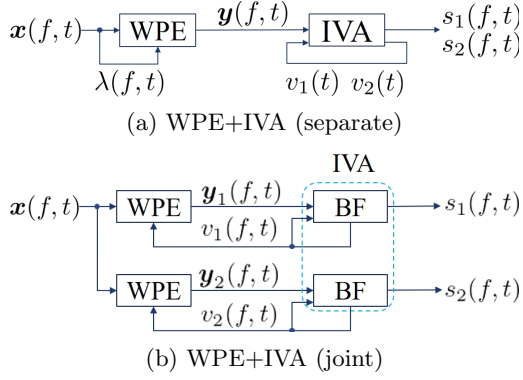


Fig. 1: WPE+IVA の分離と更新フロー. $\lambda(f, t)$ と $v_n(t)$ はそれぞれ観測信号 $\mathbf{x}(f, t)$ と音源信号 $\mathbf{s}_n(f, t)$ から求める.

4 提案手法

前節で紹介した WPE+IVA (separate) は、音源分離と残響除去で別々に最適化しているため、同時最適化が保証されていない。そこで、本節では音源分離と残響除去を同時最適化したオンライン手法 (WPE+IVA (joint)) を提案する。

4.1 分離モデルの定式化

まず、残響除去と音源分離を導出する為に観測信号 $\mathbf{x}(f, t)$ と音源信号 $\mathbf{s}(f, t)$ の関係が次の畳み込みビームフォーマで表せることを仮定する:

$$\mathbf{s}(f, t) = \mathbf{W}^H(f, 0)\mathbf{x}(f, t) + \sum_{\tau=D}^{D+L-1} \mathbf{W}^H(f, \tau)\mathbf{x}(f, t - \tau). \quad (1)$$

ただし、 $\mathbf{W}(f, \tau)$ は係数行列を表し、 $(\cdot)^H$ はエルミート転置である。 L は畳み込み演算の次数、 D は残響除去の予測遅延である。

次に、近年提案された畳み込みビームフォーマの音源別因子分解を用いる [18]。まず、式 (1) において、 $\mathbf{s}(f, t)$ を複数の畳み込みビームフォーマに分割する:

$$\mathbf{s}_n(f, t) = \mathbf{w}_n^H(f, 0)\mathbf{x}(f, t) + \sum_{\tau=D}^{D+L-1} \mathbf{w}_n^H(f, \tau)\mathbf{x}(f, t - \tau). \quad (2)$$

ただし、 $\mathbf{w}_n(f, \tau) \in \mathbb{C}^M$ ($\tau = 0, D, \dots, D + L - 1$) は n 列目の係数行列 $\mathbf{W}(f, \tau)$ である。次に、式 (2) を以下のように書き直す:

$$\mathbf{y}_n(f, t) = \mathbf{x}(f, t) - \mathbf{G}_n^H(f)\bar{\mathbf{x}}(f, t), \quad (3)$$

$$\mathbf{s}_n(f, t) = \mathbf{q}_n^H(f)\mathbf{y}_n(f, t). \quad (4)$$

WPE+IVA (joint) のフローを図 1(b) に示す。式 (3) では残響除去フィルタ $\mathbf{G}_n(f) \in \mathbb{C}^{ML \times M}$ を用いて、 n 番目の音源に含まれる残響を取り除く。ここで、 $\bar{\mathbf{x}}(f, t) = [\mathbf{x}(f, t - D)^T, \dots, \mathbf{x}(f, t - D - L + 1)^T]^T \in \mathbb{C}^{ML}$ は時刻 $t - D$ から $t - D - L + 1$ までの観測信号の系列である。式 (4) ではビームフォーマ $\mathbf{q}_n(f) \in \mathbb{C}^M$ を用いて、 n 番目の音源を抽出する。式 (2) とそれを分割した式 (3), (4) は等価の関係にある [18]。

次に、観測信号が生成されるプロセスを生成モデルにより記述する。 n 番目の音源信号の複素スペクトログラム $\mathbf{s}_n(f, t)$ を

$$\mathbf{s}_n(f, t) \sim \mathcal{N}_{\mathbb{C}}(0, v_n(t)) \quad (5)$$

のように平均が 0、分散が $v_n(t) = \mathbb{E}[|s_n(f, t)|^2]$ の複素正規分布に従う確率変数と仮定する。この仮定から、過去の時刻 t' から現在の時刻 t までの観測信号 $\mathcal{X}_t = \{\mathbf{x}_m(f, t')\}_{f, t' \leq t, n}$ が与えられたときの忘却係数付き負対数尤度関数 \mathcal{I} は以下のように記述できる:

$$\begin{aligned} \mathcal{I}(\mathcal{X}_t) \triangleq & -2 \sum_f \log |\det \mathbf{Q}(f; t)| \\ & + \frac{1}{\sum_{t' \leq t} \beta^{t-t'}} \sum_{f, t' \leq t, n} \beta^{t-t'} \left(\log v_n(t') + \frac{|s_n(f, t')|^2}{v_n(t')} \right). \end{aligned} \quad (6)$$

ただし、 $\mathbf{Q}(f) = [\mathbf{q}_1(f), \dots, \mathbf{q}_N(f)]$ である。 \triangleq はパラメータに依存しない項を除いた等号を表す。 $0 < \beta < 1$ は忘却係数である。

4.2 最適化アルゴリズム

本節では、以下に示す 3 つのパラメータ $\mathcal{G}_t = \{\mathbf{G}_n(f; t)\}_{f, n}$ 、 $\mathcal{Q}_t = \{\mathbf{Q}(f; t)\}_f$ 、 $\mathcal{V}_t = \{v_n(t)\}_n$ を再帰的に更新するオンライン最適化アルゴリズムを説明する。式 (6) は各時刻で次の更新を行う事で局所的な最適解を求めることができる:

$$\mathcal{V}_t \leftarrow \underset{\mathcal{V}_t}{\operatorname{argmin}} \mathcal{I}(\mathcal{X}_t; \mathcal{G}_{t-1}, \mathcal{Q}_{t-1}, \mathcal{V}_t), \quad (7)$$

$$\mathcal{G}_t \leftarrow \underset{\mathcal{G}_t}{\operatorname{argmin}} \mathcal{I}(\mathcal{X}_t; \mathcal{G}_t, \mathcal{Q}_{t-1}, \mathcal{V}_t), \quad (8)$$

$$\mathcal{Q}_t \leftarrow \underset{\mathcal{Q}_t}{\operatorname{argmin}} \mathcal{I}(\mathcal{X}_t; \mathcal{G}_t, \mathcal{Q}_t, \mathcal{V}_t). \quad (9)$$

各変数を最適化するオンラインアルゴリズムを Algorithm 1 にまとめた。ここで、 \mathbf{I}_M を $M \times M$ の単位行列、 $\mathbf{0}$ を零行列とする。以降、 \mathcal{G}_t 、 \mathcal{Q}_t 、 \mathcal{V}_t の更新式について簡潔に述べる。

\mathcal{V}_t の更新式を Algorithm 2 に示した。IVA では、各時刻で得られる分離信号のパワーの平均を取ることと式 (6) の最適化と各周波数ビンで得られる分離信号の順序整理を同時に行う。

次に、 \mathcal{G}_t の更新式を求める。 \mathcal{V}_t を固定すると式 (6) は分離フィルタ \mathcal{Q}_{t-1} に依存せず次のように求めることで最小化できる [18]:

$$\mathbf{G}_n(f; t) = \mathbf{R}_n^{-1}(f; t)\mathbf{P}_n(f; t). \quad (10)$$

ただし、 $\mathbf{R}_n(f; t) = \beta \mathbf{R}_n(f; t - 1) + \frac{\bar{\mathbf{x}}(f, t)\bar{\mathbf{x}}^H(f, t)}{v_n(t)}$ と $\mathbf{P}_n(f; t) = \beta \mathbf{P}_n(f; t - 1) + \frac{\bar{\mathbf{x}}(f, t)\mathbf{x}^H(f, t)}{v_n(t)}$ は共分散行列である。逆行列定理 [21] を用いることで、残響除去フィルタ $\mathbf{G}_n(f; t)$ は Algorithm 3 のように再帰的に求められる。ただし、 $\mathbf{K}_n(f, t)$ はカルマンゲインである。この更新式は分散 $v_n(t)$ を観測信号 $\mathbf{x}(f, t)$ からではなく分離信号 $\mathbf{s}(f, t)$ から求める点を除くと従来のオンライン WPE [15] と同じである。また、従来の同時最適化の更新式 [17] と比べると推定する残響除去フィルタ $\mathbf{G}_n(f; t)$ は 1 個から M 個に増えるが、1 個の残響除去フィルタで計算する $\mathbf{R}_n(f; t)$ の行列の行数が M^2L から ML に小さくなり、逆行列演算量は行数の 3 乗の計算オーダーであるため、オンライン化に重要な計算コストを大幅に削減できている。

最後に、 \mathcal{Q}_t の更新式を示す。残響除去フィルタ \mathcal{G}_t を固定すると、従来の IP 法 [7, 8] と同じ方法で \mathcal{Q}_t を更新することができる。ただし、更新に用いる共分散行列 $\Sigma_n(f, t)$ は観測信号 $\mathbf{x}(f, t)$ ではなく残響除去後の信号 $\mathbf{y}_n(f, t)$ から求める。本稿では $\Sigma_n(f, t)$ と $\mathbf{q}_n(f; t)$ を [9] に習い、Algorithm 4 のように再帰的

Algorithm 1 Update rule for WPE+IVA (joint)

```
1: Initialization:  $\mathbf{Q}_n(f; 0) = \mathbf{I}_M$ ,  $\mathbf{R}_n^{-1}(f; 0) = \mathbf{I}_{ML}$ ,  $\mathbf{G}_n(f; 0) = \mathbf{0}$ , and  $\Sigma_n(f, 0) = 10^{-5} \times \mathbf{I}_M$ .
2: for  $t = 1$  to  $T$  do
3:    $\mathcal{G}_{t-1}$ ,  $\mathcal{Q}_{t-1}$  を用いて式 (3), (4) を計算する.
4:   for  $n = 1$  to  $N$  do
5:     (Update  $\mathcal{V}_t$ )
6:      $\mathbf{Q}(f, t) = \mathbf{Q}(f, t-1)$ 
7:     for  $f = 1$  to  $F$  do
8:       (Update  $\mathcal{G}_t$ )
9:       (Update  $\mathcal{Q}_t$ )
10:    end for
11:  end for
12: end for
```

Algorithm 2 Update \mathcal{V}_t

```
1:  $v_n(t) \leftarrow \sum_{f=1}^F |s_n(f, t)|^2 / F$ 
```

に求めた。ただし、 $0 < \alpha < 1$ は β とは独立した忘却係数であり、 L_b はブロック長である。 \mathbf{e}_n は単位行列 \mathbf{I}_M の第 n 列のベクトルである。

5 関連研究

近年、周波数領域で推定した分離フィルタを時間領域の FIR フィルタに変換し、非因果成分の一部を残してカットすることで低遅延でオンライン動作する BSS 手法が提案されている [22]。この手法は補聴器を想定した音源分離実験により、分離性能を維持しつつアルゴリズム遅延を 10 ms まで短くできることが確認されている。

それに対し、本提案手法は残響除去を行うため、残響を取り除いた明瞭度の高い音源分離が行える点で異なる。両方の手法の比較や統合した手法の研究は今後の研究課題とする。

6 評価実験

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 6 話者、女性 4 話者計 10 話者分のデータを使用した [23]。このデータセットの中からランダムに選択した 2 人の異なる話者の発話に対し、実測の車室内インパルス応答を畳み込むことで、2 チャンネル、信号長 20 秒の観測信号を 100 パターン作成した。サンプリング周波数は 16 kHz とした。インパルス応答は車室内で録音した時間引き延ばしパルス (Time-Stretched Pulse: TSP) を用いて測定した。音源を自動車の運転席と助手席に配置し、マイクロホンを車室内前方のマップランプに取り付けて TSP 信号を録音した。音源とマイクロホンの配置を図 2 に示す。車室内の残響時間 (T_{60}) は 58 ms であった。本稿では、雑音を含む場合と含まない場合の両方を想定した実験を行った。雑音は、一般道を走行中の自動車から収録して、観測信号に信号対雑音比 (Signal-to-Noise Ratio: SNR [24]) 0 dB もしくは 10 dB で足した。実験では、ベースライン手法で述べたオンライン IVA、WPE+IVA (separate) と提案手法である WPE+IVA (joint) の 3 つの手法を比較した。分離性能の評価には、分離信号の品質 (Signal-to-Distortion Ratio: SDR)、分離信号と音源信号との分

Algorithm 3 Update \mathcal{G}_t

```
1:  $\mathbf{K}_n(f, t) \leftarrow \frac{\mathbf{R}_n^{-1}(f; t-1)\bar{\mathbf{x}}(f, t)}{\beta v_n(t) + \bar{\mathbf{x}}^H(f, t)\mathbf{R}_n^{-1}(f; t-1)\bar{\mathbf{x}}(f, t)}$ 
2:  $\mathbf{R}_n^{-1}(f; t) \leftarrow \frac{\mathbf{R}_n^{-1}(f; t-1) - \mathbf{K}_n(f, t)\bar{\mathbf{x}}^H(f, t)\mathbf{R}_n^{-1}(f; t-1)}{\beta}$ 
3:  $\mathbf{G}_n(f; t) \leftarrow \mathbf{G}_n(f; t-1) + \mathbf{K}_n(f, t)\mathbf{y}^H(f, t)$ 
```

Algorithm 4 Update \mathcal{Q}_t

```
1:  $\Sigma_n(f, t) \leftarrow \alpha \Sigma_n(f, t - L_b)$   
    $+ (1 - \alpha) \cdot \frac{1}{L_b} \sum_{\tau=t-L_b+1}^t \frac{\mathbf{y}_n(f, \tau)\mathbf{y}_n^H(f, \tau)}{v_n(\tau)}$ 
2:  $\mathbf{q}_n(f; t) \leftarrow (\mathbf{Q}^H(f; t)\Sigma_n(f, t))^{-1}\mathbf{e}_n$ 
3:  $\mathbf{q}_n(f; t) \leftarrow \frac{\mathbf{q}_n(f; t)}{\sqrt{\mathbf{q}_n^H(f; t)\Sigma_n(f, t)\mathbf{q}_n(f; t)}}$ 
```

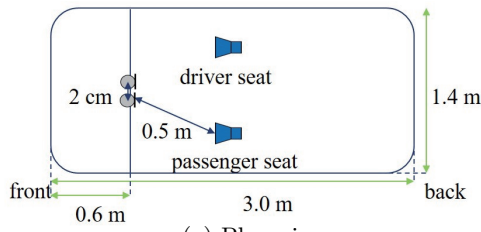
離度合 (Signal-to-Interferences Ratio: SIR)、分離信号に含まれる雑音と音源信号との SNR を用いる [24]。SDR, SIR, SNR は値が高い程高い分離性能を示す。アルゴリズム遅延と計算効率を考慮し、STFT にハン窓を使用し、窓長と移動長は 8, 4 ms とした。その他音源分離のパラメータは次のように定めた $\{\alpha = 0.96, \beta = 0.9999, L_b = 2, D = 1, L = 5\}$ 。

図 3 は、各々の音源分離手法の 2 秒毎の SIR の平均改善量を示している。図より、雑音の有無にかかわらず、提案手法の SIR の平均改善量がベースライン手法を上回ることを確認した。特に、図 3(c) の場合では、WPE+IVA (separate) の SIR の平均改善量がオンライン IVA と変わらない一方で、提案手法はオンライン IVA より改善していることを確認した。表 1 は、各音源分離手法での 20 秒全体での SDR, SIR, SNR の平均改善量を示している。表より、雑音を含まない場合では提案手法の SDR の平均改善量がベースライン手法を上回らない一方で、雑音を含む場合では上回ることを確認した。SIR, SNR の改善量も提案手法がベースライン手法を上回ることも確認した。特に、SNR 0 dB の場合では、WPE+IVA (separate) の SIR の平均改善量がオンライン IVA より劣る一方で、提案手法は上回ることを確認した。以上の実験結果から、提案手法が観測信号ではなく分離信号の分散を用いて残響除去フィルタを更新し、音源と雑音をより区別して分離したことで SDR, SIR 及び SNR の平均改善量が向上したと考察される。

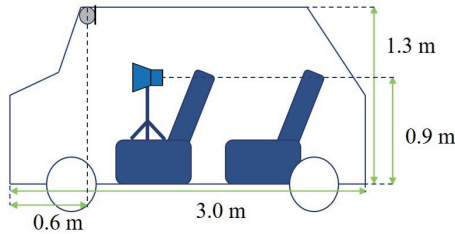
最後に、提案手法の計算効率性を評価した。Intel Core 3.6 GHz 1-core CPU が導入された PC 上で Python 3.7.7 で実装したシステムを用いて、信号長 20 秒の観測信号を分離する時間を測定した。表 2 は、2 チャンネル、4 チャンネルの観測信号を分離する計算時間と想定される遅延の総和を示している。表より、[4] で示される快適な車室内コミュニケーションを実現する為の遅延の上限 12 ミリ秒より短い遅延で分離できており、提案手法は実時間動作を十分に満たしていることを確認した。

7 結論

本稿では、低遅延でオンライン動作するブラインド音源分離手法を提案した。提案手法はオンラインで残響除去と音源分離を同時に行う事で従来手法の単純結合より高い分離性能を実現できる。車室内環境下を想定した音源分離実験により、雑音を含む短い残響環境下での提案手法が比較的小さいアルゴリズム遅延 (= 8 ms) と計算遅延 (= 0.98 ms) で分離性能が向上することを確かめた。



(a) Plan view



(b) Side view

Fig. 2: 音源とマイクロホンの配置

Table 1: 各手法での平均 SDR, SIR, SNR 改善量 [dB] の比較. ボールド体は各評価指標で最も高い値を表す.

Method (雑音無し)	SDR	SIR	SNR
オンライン IVA	9.86	16.79	-
WPE+IVA (separate)	11.59	20.54	-
WPE+IVA (joint)	9.66	22.35	-
Method (SNR = 10[dB])	SDR	SIR	SNR
オンライン IVA	6.52	15.84	-3.63
WPE+IVA (separate)	6.50	16.41	-3.82
WPE+IVA (joint)	7.85	18.69	-2.78
Method (SNR = 0[dB])	SDR	SIR	SNR
オンライン IVA	3.54	14.61	-1.68
WPE+IVA (separate)	3.81	14.38	-1.44
WPE+IVA (joint)	6.81	15.79	1.38

参考文献

[1] J. Antoni, *Mechanical Systems and Signal Processing*, vol. 19, no. 6, pp. 1166–1180, 2005.

[2] M. Zoulikha *et al.*, *Applied Acoustics*, vol. 112, pp. 192–200, 2016.

[3] R. Landgraf *et al.*, in *Proc. Speech Prosody*, 2016, pp. 479–483.

[4] A. Theiss *et al.*, in *Speech Communication; 11. ITG Symposium*. VDE, 2014, pp. 1–4.

[5] A. Hyvärinen *et al.*, *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[6] T. Kim *et al.*, in *Proc. ICA*, 2006, pp. 165–172.

[7] N. Ono, in *Proc. WASPAA*, 2011, pp. 189–192.

[8] N. Ono *et al.*, in *Proc. LVA/ICA*. Springer, 2010, pp. 165–172.

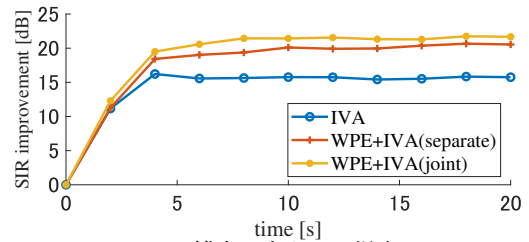
[9] T. Taniguchi *et al.*, in *Proc. HSCMA*, 2014, pp. 107–111.

[10] D. Mauler *et al.*, in *Proc. EUSIPICO*, 2007, pp. 222–226.

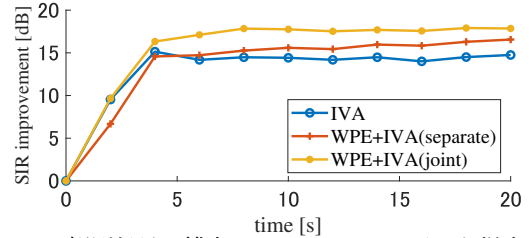
[11] M. Stone *et al.*, *Ear and Hearing*, vol. 29, no. 4, pp. 601–617, 2008.

[12] S. Araki *et al.*, *IEEE Trans. ASLP*, vol. 11, no. 2, pp. 109–116, 2003.

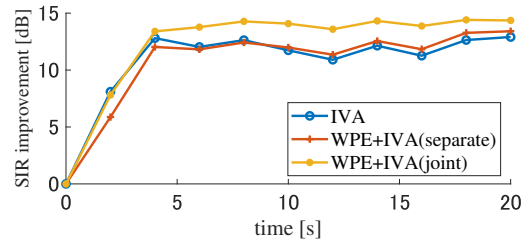
[13] T. Nakatani *et al.*, in *Proc. ICASSP*, 2008, pp. 85–88.



(a) 雑音を含まない場合



(b) 観測信号に雑音を SNR 10 dB で足した場合



(c) 観測信号に雑音を SNR 0 dB で足した場合

Fig. 3: 各手法での 2 秒毎の SIR の平均改善量の比較

Table 2: 20 秒 (≒ 5000 フレーム) の観測信号を分離する計算時間と想定される遅延の総和.

各時間 [ms]	2ch	4ch
全体の計算時間	2.7×10^3	4.9×10^3
1 フレーム当たりの計算遅延	0.54	0.98
アルゴリズム遅延	8.0	
想定される遅延の総和	8.54	8.98
許容できる遅延の上限 [4]	12.0	

[14] T. Yoshioka *et al.*, in *Proc. ICASSP*, 2009, pp. 3733–3736.

[15] J. Caroselli *et al.*, in *Proc. Interspeech*, 2017, pp. 3877–3881.

[16] T. Yoshioka *et al.*, *IEEE Trans. ASLP*, vol. 19, no. 1, pp. 69–84, 2010.

[17] H. Kagami *et al.*, in *Proc. ICASSP*, 2018, pp. 31–35.

[18] T. Nakatani *et al.*, *IEEE/ACM Trans. ASLP*, vol. 28, pp. 2267–2282, 2020.

[19] T. Nakatani *et al.*, in *Proc. Interspeech*, 2020.

[20] T. Ueda *et al.*, “Low latency online blind source separation based on joint optimization with blind dereverberation,” in *Proc. ICASSP*, 2021 (submitting).

[21] S. Haykin, Pearson Education India, 2008.

[22] M. Sunohara *et al.*, in *Proc. ICASSP*, 2017, pp. 216–220.

[23] A. Kurematsu *et al.*, *Speech communication*, vol. 9, no. 4, pp. 357–363, 1990.

[24] E. Vincent *et al.*, *IEEE/ACM Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.