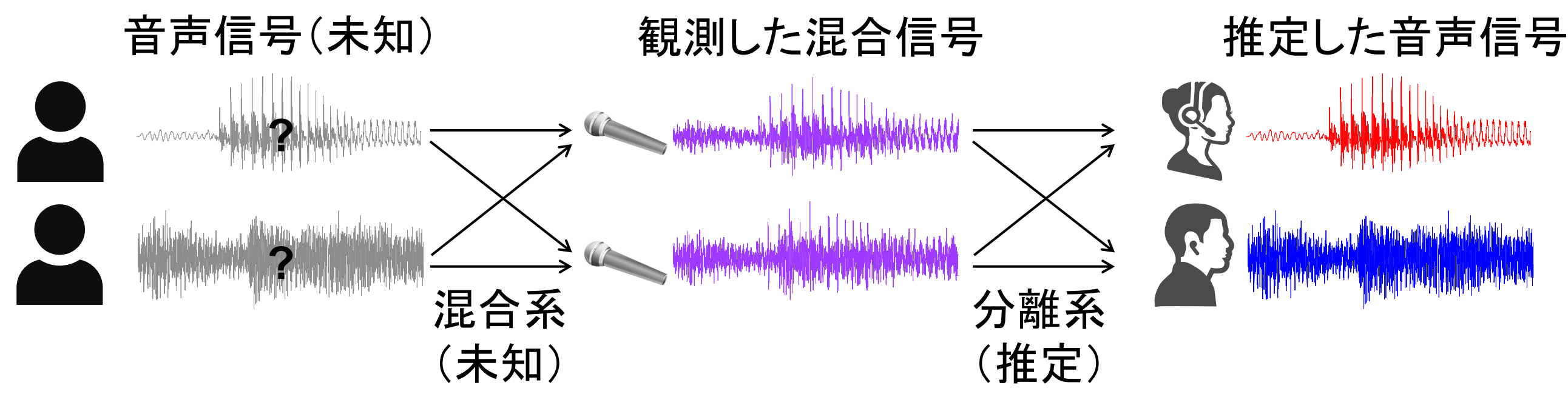


1. 研究背景

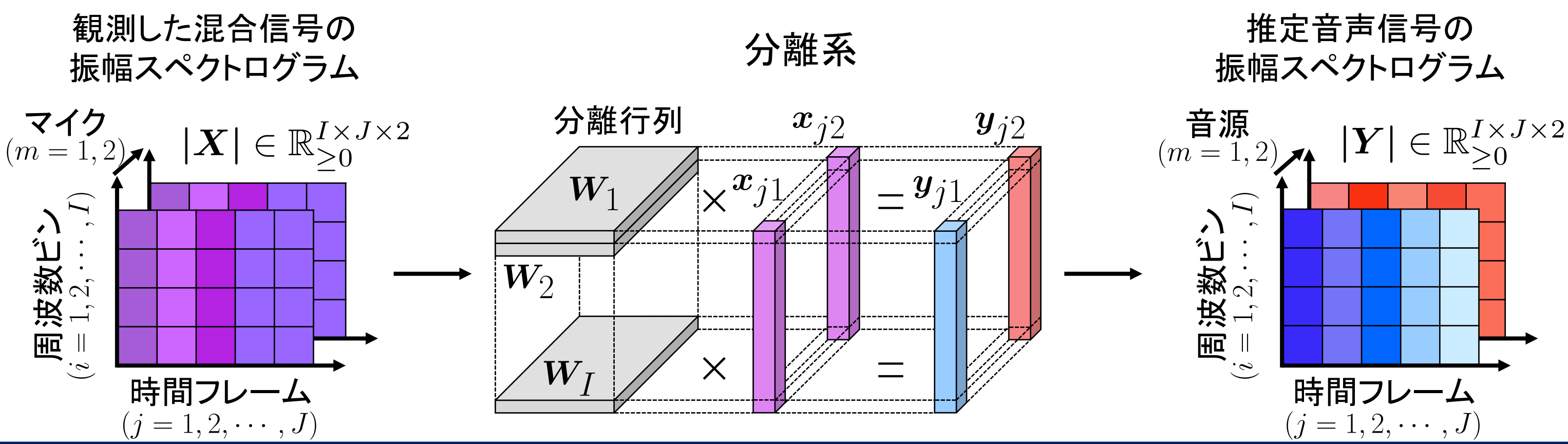
ブラインド音源分離 (blind source separation: BSS)

- 複数の音源が混ざり合った信号から各音源の信号を推定



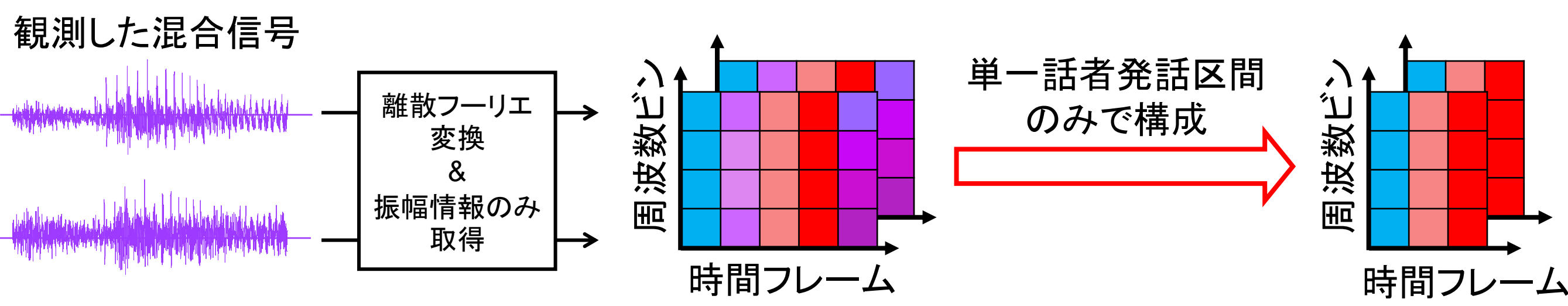
独立ベクトル分析 (independent vector analysis: IVA) [Kim+, 2007]

- BSS手法の一つであり, 本研究では振幅スペクトログラムの入力を仮定



本研究の目的

- IVAは**単一話者発話区間**だけを観測信号とした場合に最良の性能を達成できることが理論的に証明されている [Jianjun+, 2023]
- 論文に基づき, **観測信号から単一話者発話区間を特定し, それのみを入力として使用**することができればBSS性能を大幅に向上させることが可能となる [Kato+, 2023]



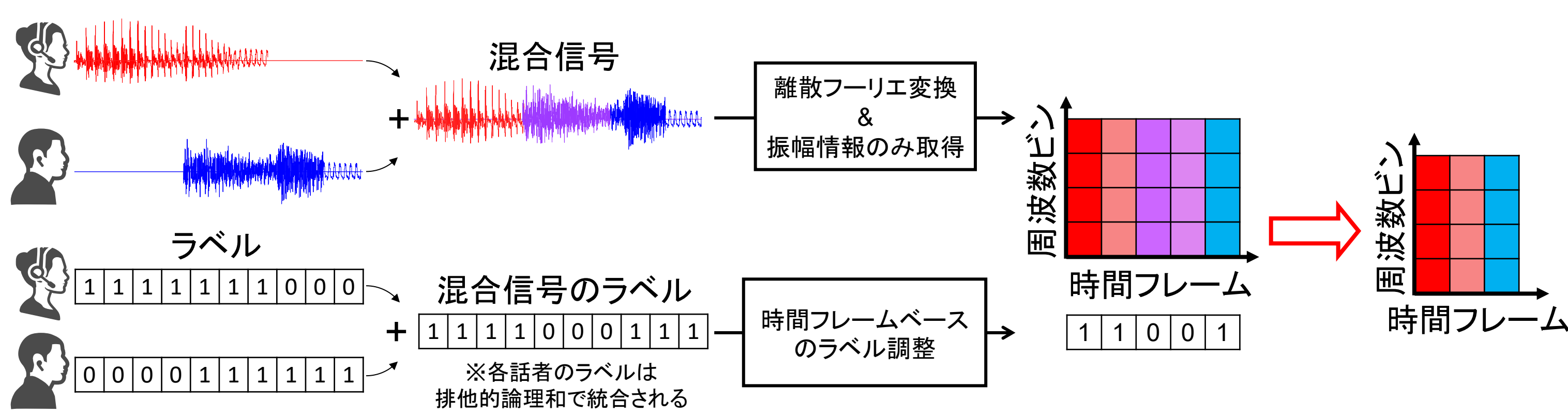
- 本研究では, 観測信号中の単一話者発話区間の締める割合によってどのようにIVAの性能が変わるかを実験的・網羅的に調査

2. 既存手法

提案手法では単一話者発話区間の抽出が必要 [Kato+, 2023]

- 本研究ではラベルを定義し, 単一話者発話区間を表現する

- 単一話者発話区間 : 1のラベル
- その他(複数話者や雑音) : 0のラベル



F-WDO (frame-level W-disjoint orthogonality) [Jianjun+, 2023]

- 各時間フレームにおいて, 一つの音源が他の音源よりも支配的であることを示す概念
- 論文ではBSSのためのIVAアルゴリズムの性能に対する音源のスパース性の影響を数学的に導出し調査している。

IVAの目的関数を定式化(パーミュテーションがない場合)

$$J_1(W, \emptyset) = \mathbb{E} \left\{ \sum_{k=1}^K \left(\sum_{f=1}^{N_f} |s_k(f, t)|^2 \right)^{\beta/2} \right\}$$

IVAの目的関数を定式化(パーミュテーションがある場合)

$$J_1(W, \mathbf{F}_p) = \mathbb{E} \left\{ \sum_{k=1}^K \left(\sum_{f \notin \mathbf{F}_p} |s_k(f, t)|^2 + \sum_{f \in \mathbf{F}_p} |s_{\phi(k, f)}(f, t)|^2 \right)^{\beta/2} \right\}$$

パーミュテーション問題を完全に回避するための条件(分離性能向上の条件)

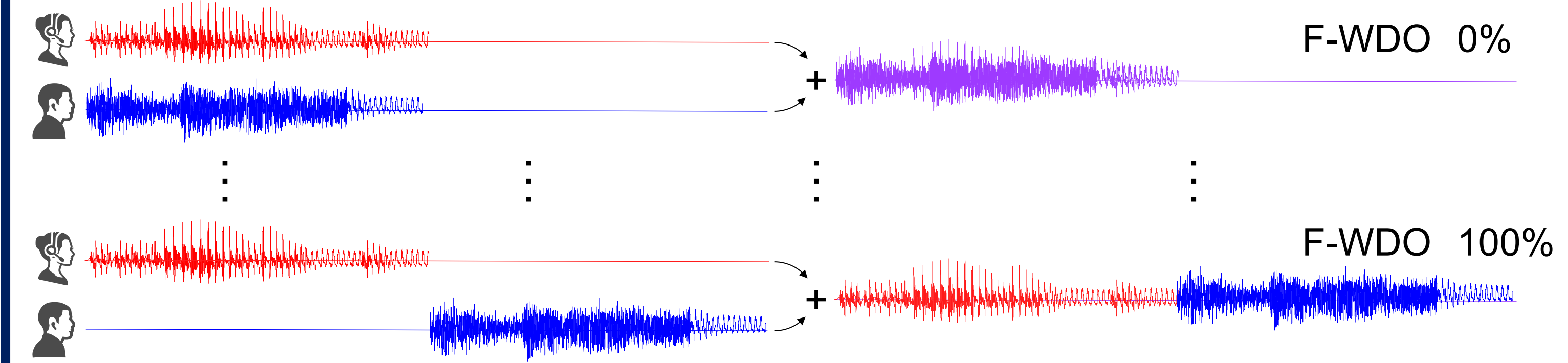
$$J_1(W, \emptyset) < J_1(W, \mathbf{F}_p), \forall \mathbf{F}_p \subseteq \{1, \dots, N_f\}, \mathbf{F}_p \neq \emptyset$$

- この条件は, 時間フレームにおいて一つの音源のエネルギーが支配的であることを要求しており, F-WDOの定義と直接的に関連
- 論文の結果では, F-WDOの度合いが増加するにつれて, 目的関数の値が減少(分離性能が向上)することが示されている

3. 実験

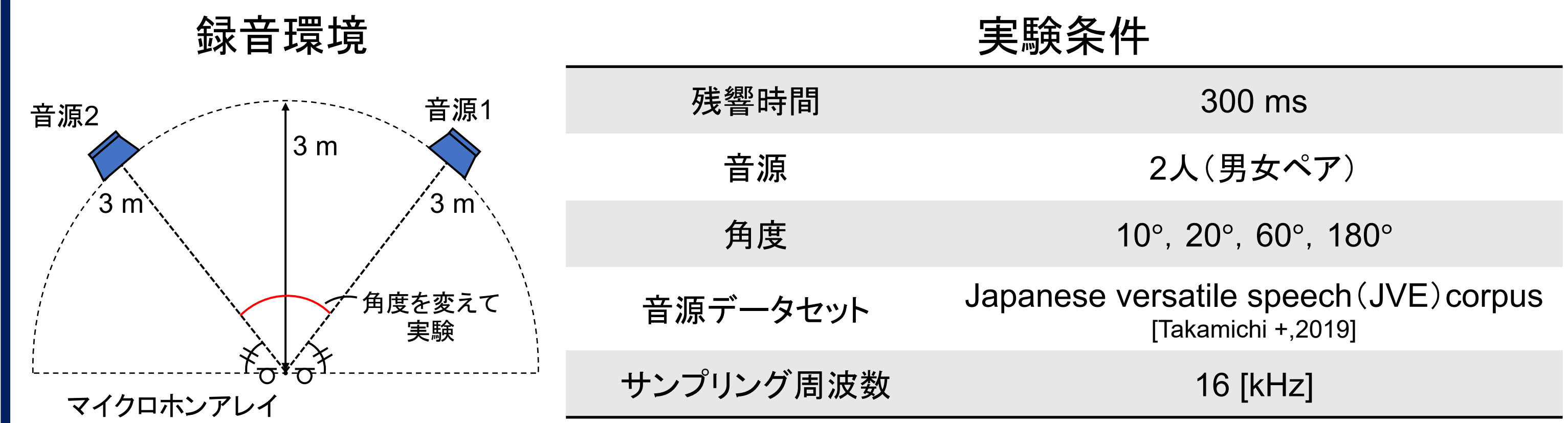
F-WDOを考慮した混合信号の作成

- 一方の話者信号を固定し, 残りの話者信号を時間シフトさせ混合信号を作成する。
- 一方の話者の時間遅延を調整することで, F-WDOの度合いを疑似的に0~100%の範囲で変化させる。



PyRoomAcousticsで部屋の録音環境シミュレーション

- Image法でインパルス応答を取得し, 各話者信号に適応することで実際の録音環境を模倣した混合信号を生成する



4. 実験結果

F-WDOが0.3~0.8の区間

- 提案手法が優位であり, 性能向上は+0.1~1 dBの範囲で観察された

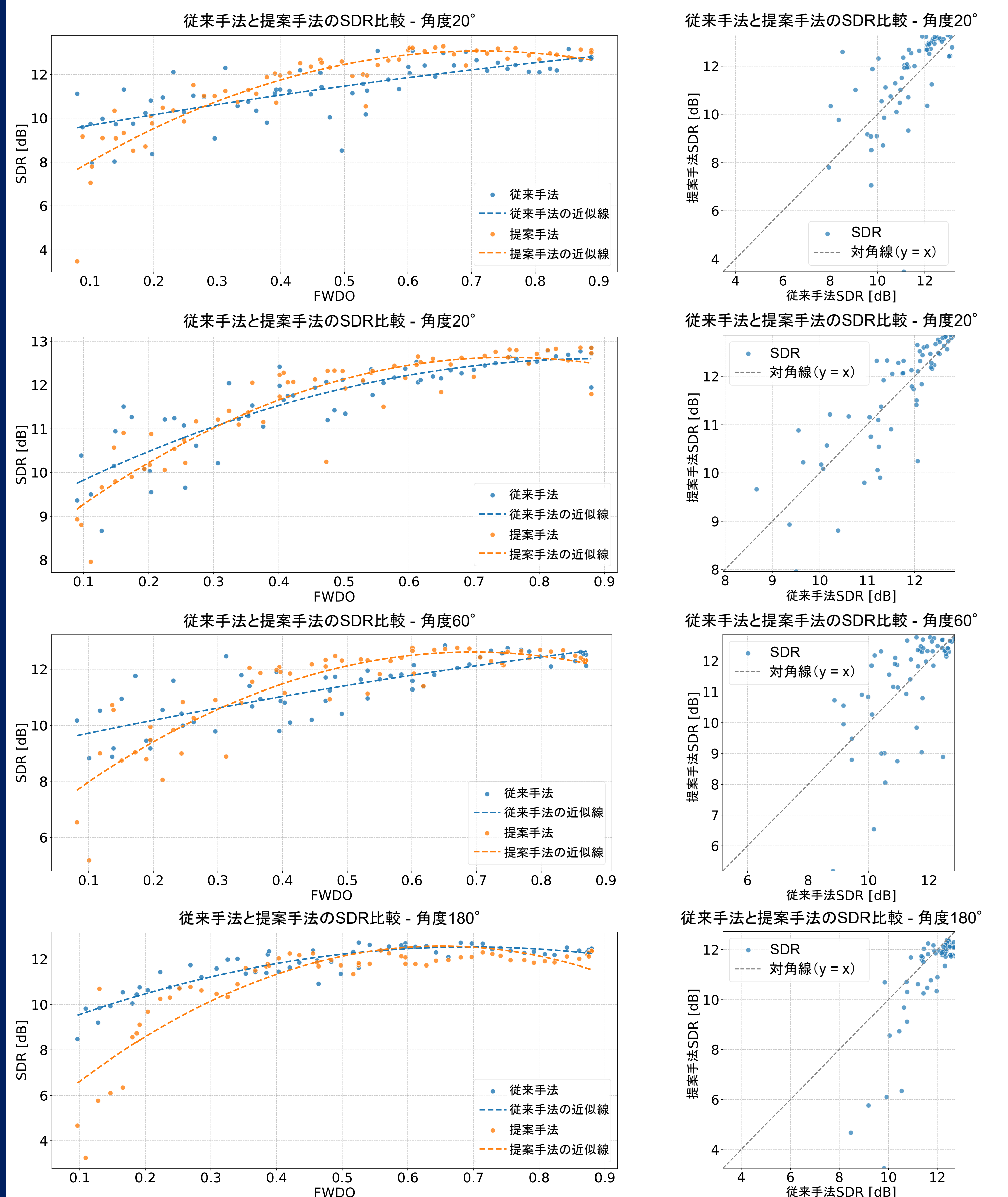
F-WDOが0.8以上

- 両手法(提案手法と従来手法)で同等の性能が得られた

F-WDOが0.3以下

- 提案手法がやや劣化し, 性能が低下した。これは, 混合信号の30%未満の情報のみを使用していることが原因と考えられる

- 条件により上記結果が得られない場合もあった。実験環境や音源の特性, 混合信号の構成などが影響している可能性がある



本研究のまとめ

- 単一話者発話区間を援用した分離行列推定法を提案し, 結果では, **提案手法が従来手法を上回る場合と下回る場合が混在した**
- この結果は, 分離精度向上が単純な要因だけでなく, **複数の要素が絡み合う複雑な問題**であることを示唆している