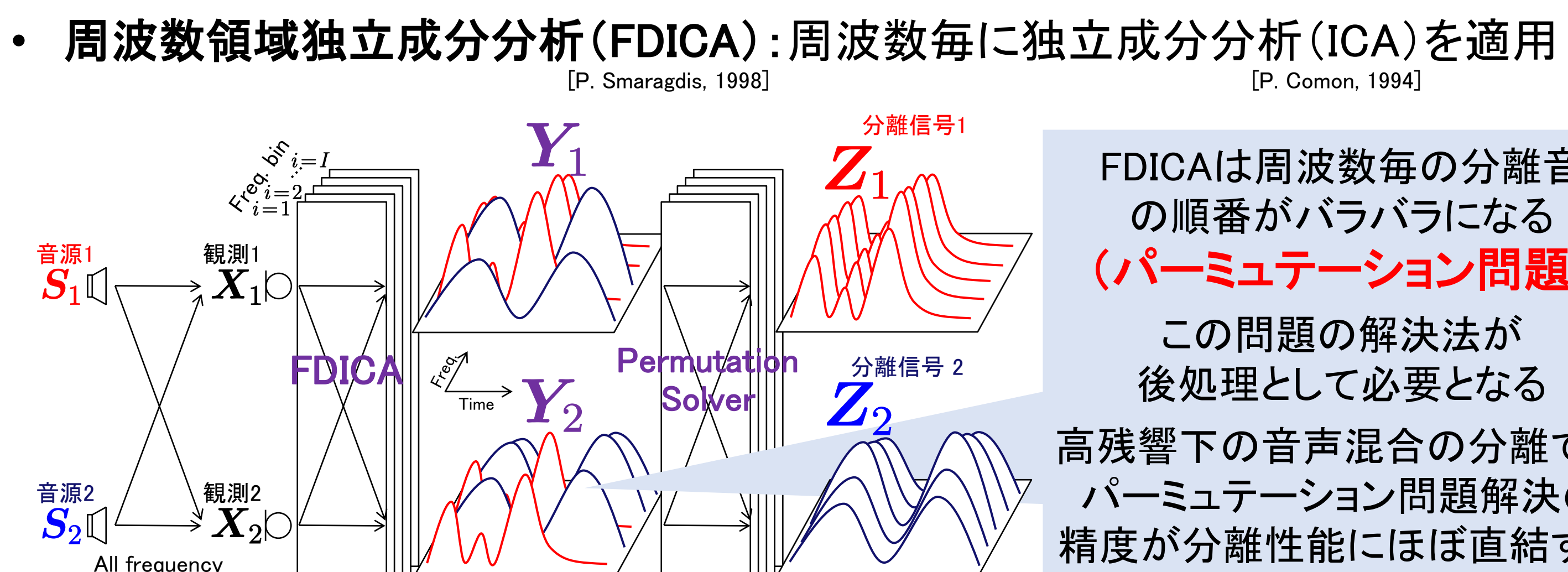
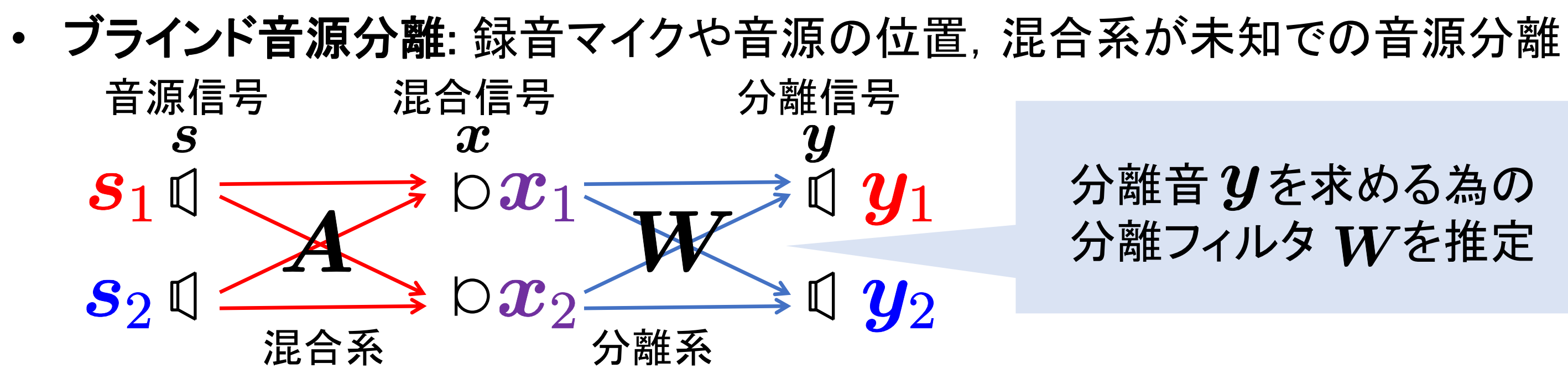


局所時間周波数構造に基づく 深層パーミュテーション解決法の実験的評価

山地修平, 北村大地 (香川高等専門学校)

1. 研究背景



研究の動機

高残響下での音声混合信号に対して, FDICAを適用した際の**周波数毎の分離結果**を確認すると, 非常に高い精度で分離されている [D. Kitamura+, 2017]

独立ベクトル分析 (IVA) [T. Kim+, 2007] や独立低ランク行列分析 (ILRMA) [D. Kitamura+, 2016] では, **単にパーミュテーション問題解決に失敗している可能性が高い**

本研究の目的

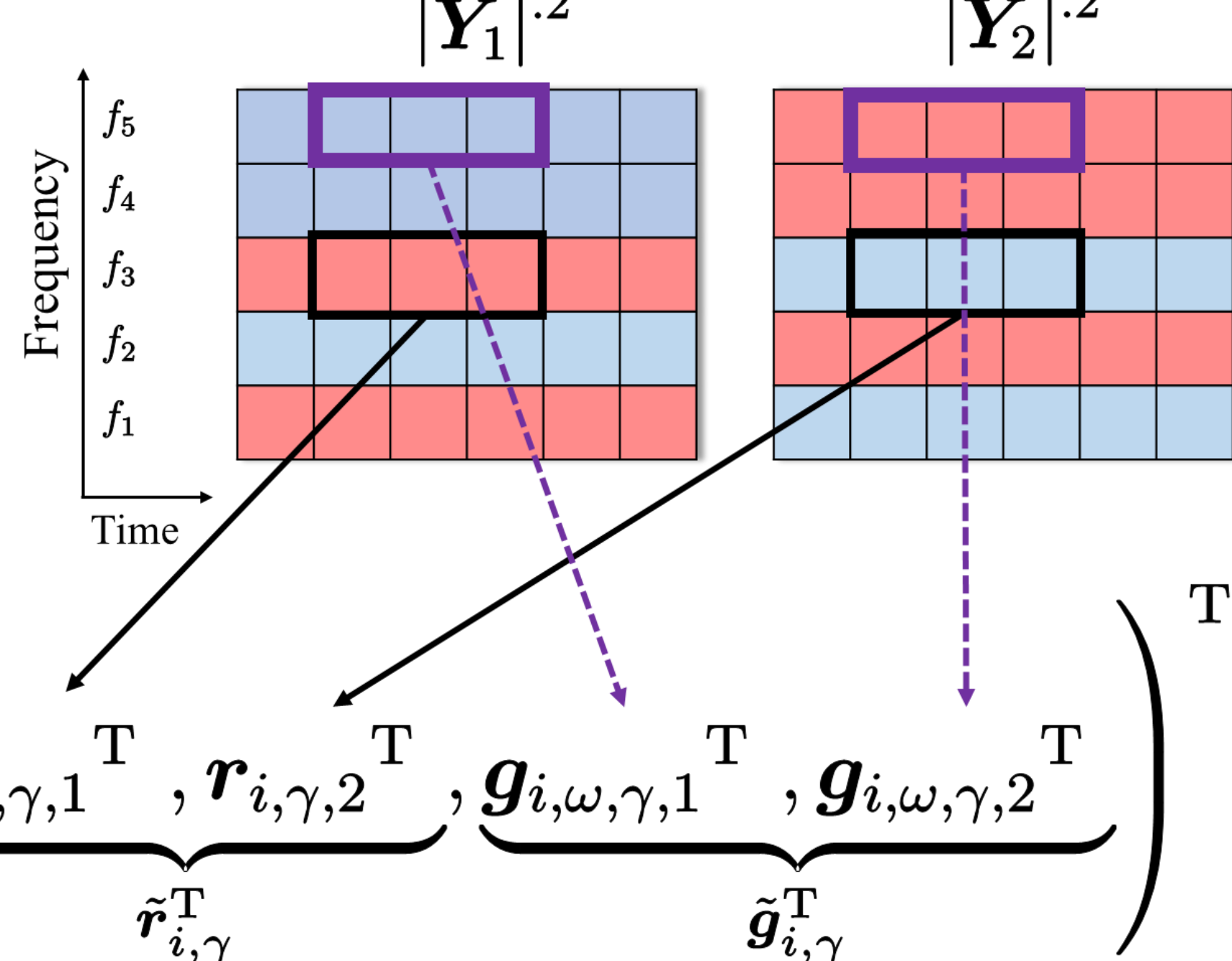
Deep neural network (DNN) に基づくパーミュテーション解決法を提案
推定結果を**時間・周波数方向に多数決**することで推定の頑健性を確保

2. DNNの入力と出力

• DNNに基づく分離信号のパーミュテーションの推定 [山地ら, 2020]

DNN に入力された**2つの周波数時系列パワー**が同一音源か否かを推定

FDICAによる推定分離信号の
パワースペクトログラム



DNNの入力

各分離信号から
**2つの周波数時系列
パワー**を抽出

入力ベクトルを正規化

$$\tilde{d}_{i,\omega,\gamma} = \frac{d_{i,\omega,\gamma}}{\|d_{i,\omega,\gamma}\|_2} \in \mathbb{R}_{\geq 0}^{4 \times 1}$$

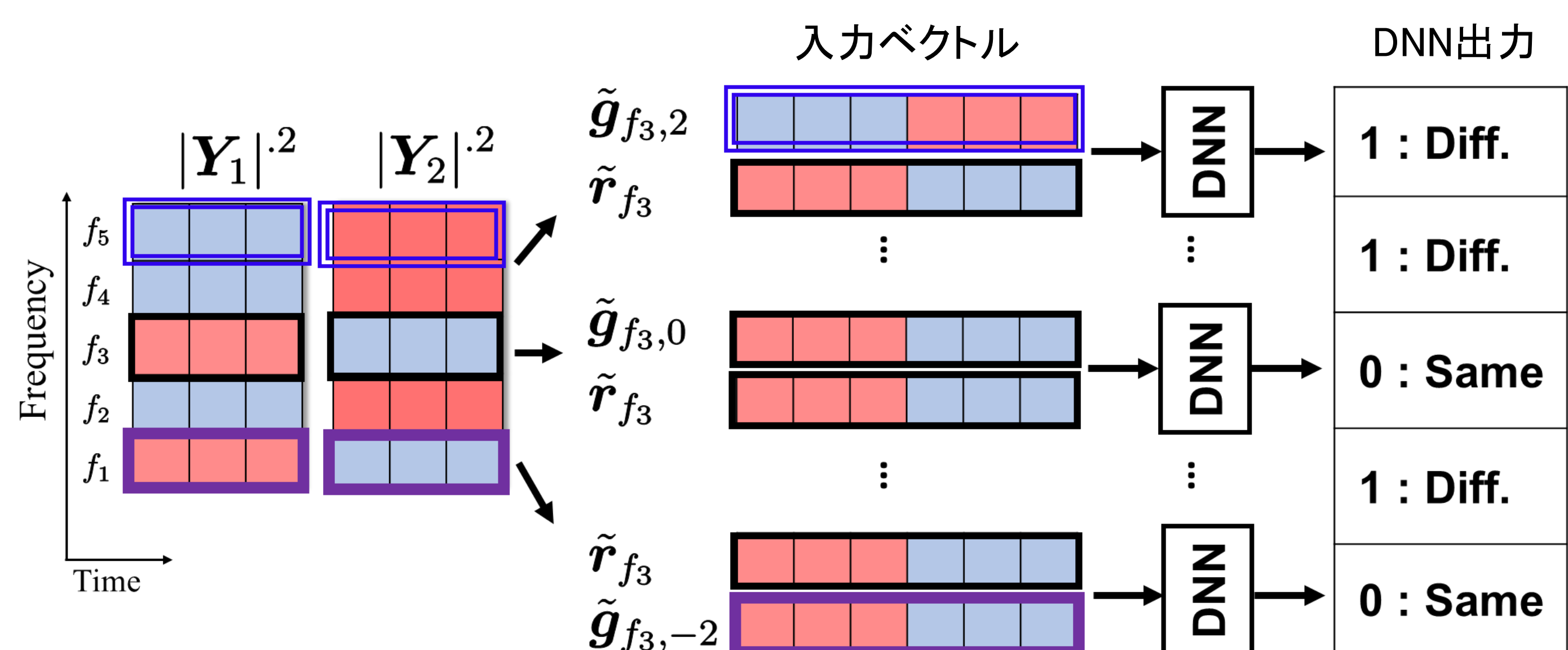
DNNの出力

$(r_{i,\gamma,1} \ \& \ g_{i,\omega,\gamma,1})$
 $(r_{i,\gamma,2} \ \& \ g_{i,\omega,\gamma,2})$

同一音源なら「0」
異なる音源なら「1」

• サブバンド周波数帯域におけるDNN推定

2本の短時間時系列パワー ($\tilde{r}_i, \tilde{g}_{i,\omega}$) の全組み合わせから
サブバンド周波数帯域内でパーミュテーション推定

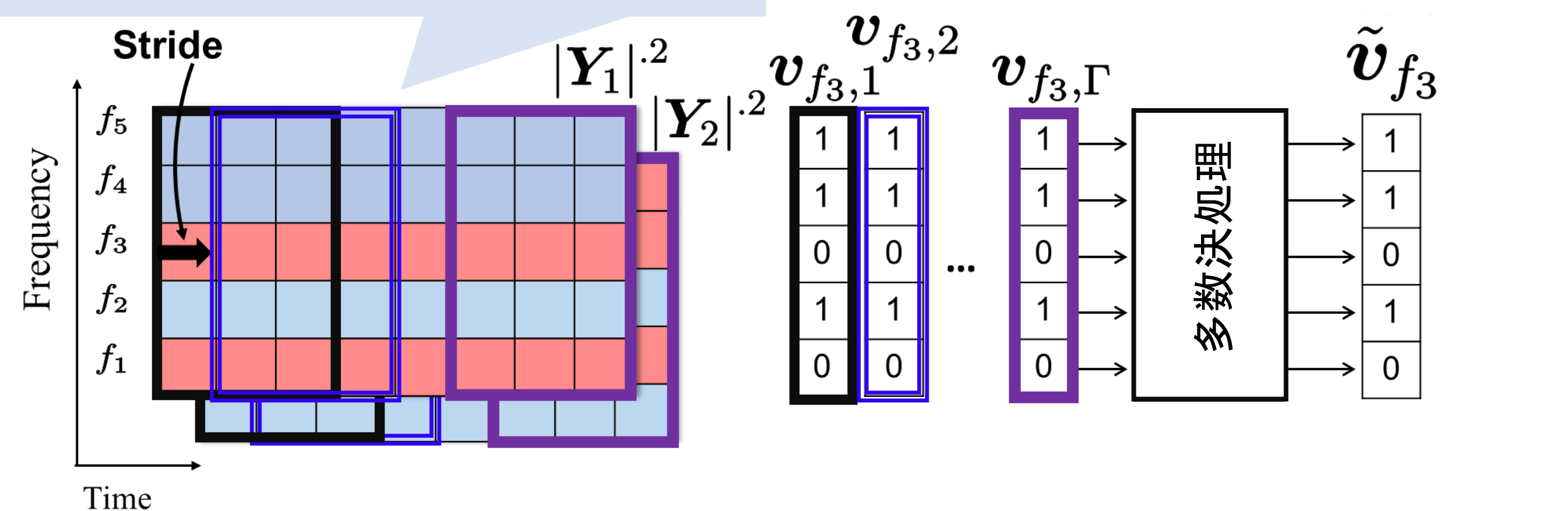


3. 時間及び周波数方向の多数決処理

• 時間方向の多数決処理

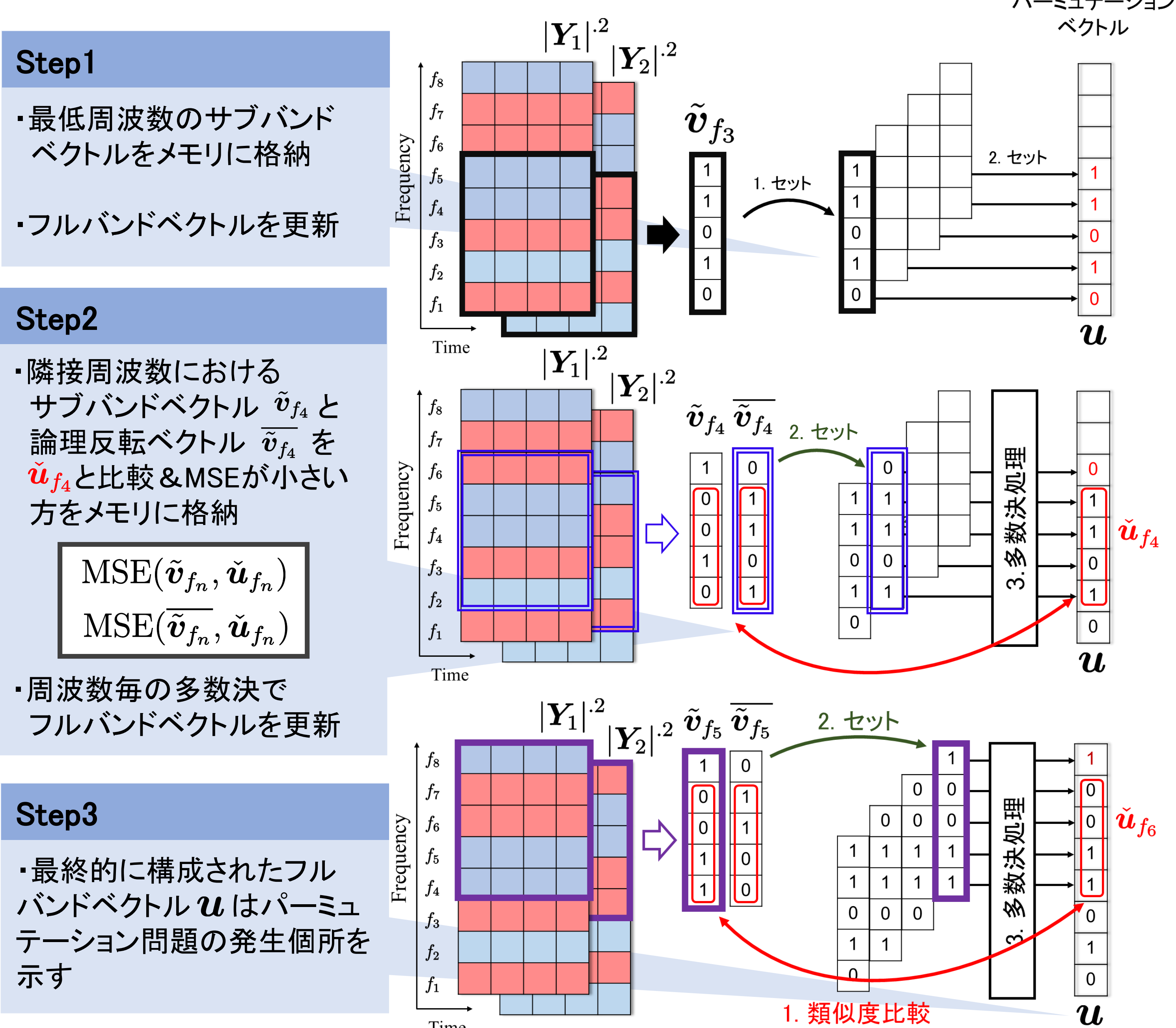
無音区間における不安定なDNN推定の影響を軽減

入力ベクトルの選択範囲を
時間方向にシフトすることで
全時間フレームにDNN推定を走査



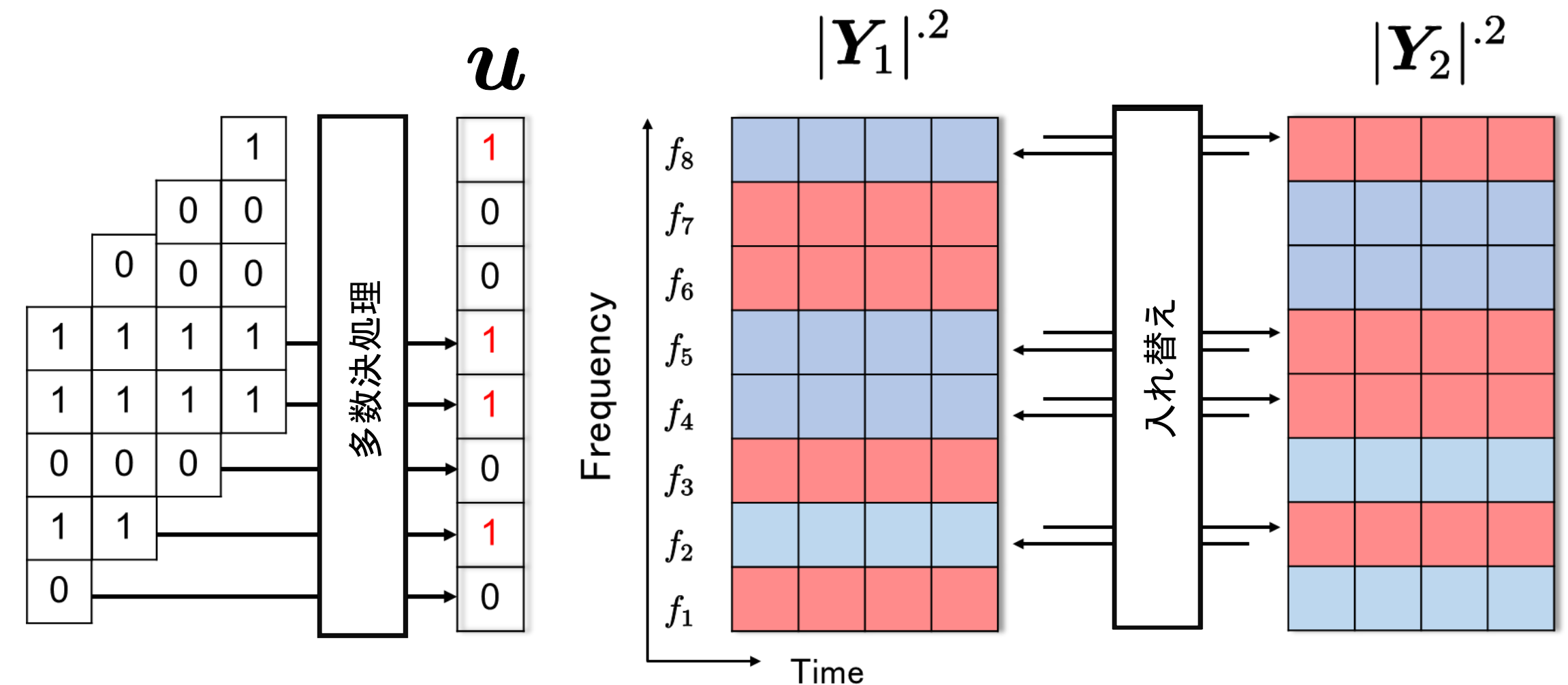
• 周波数方向の多数決処理

DNNの推定結果を全周波数で**0→音源赤, 1→音源青**に統一



• フルバンドベクトルに基づく成分の交換

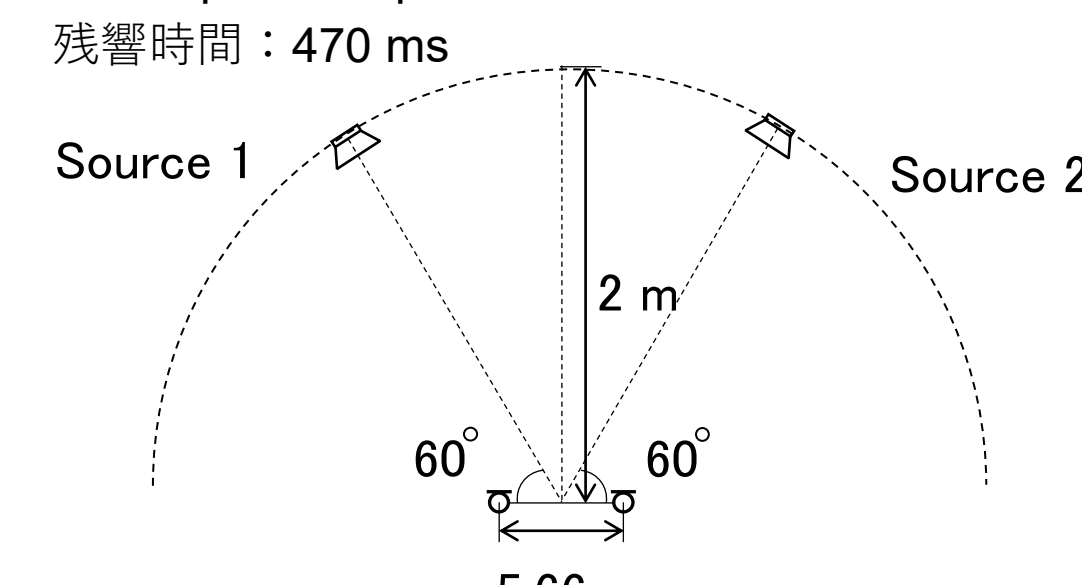
u に基づく分離信号成分の入れ替えによるパーミュテーション解決



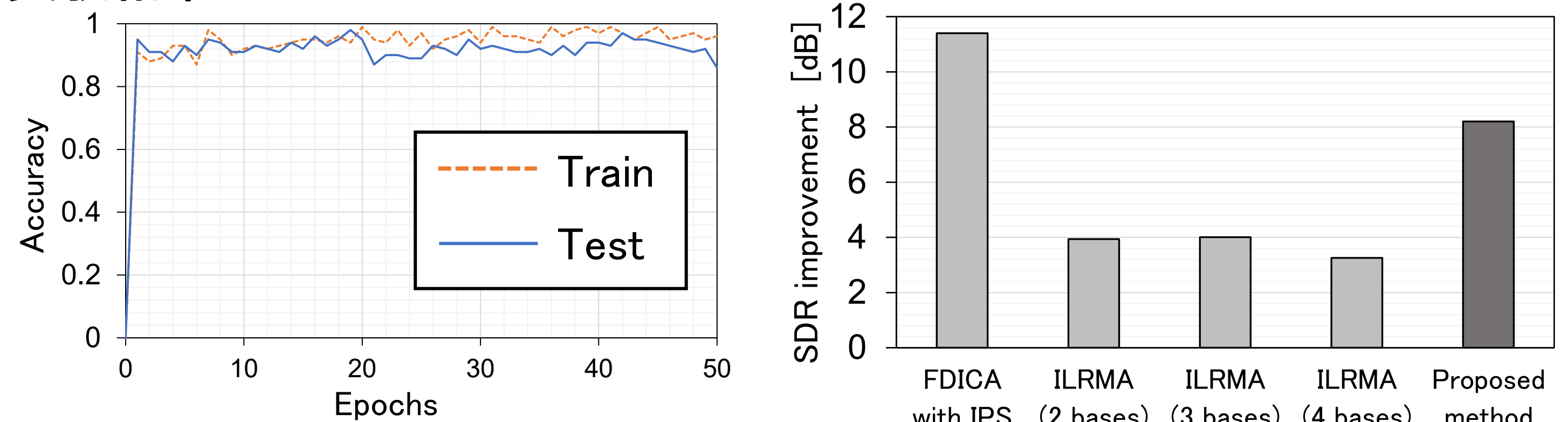
4. 比較実験

• 実験条件

学習用音声信号	JVSコーパスの音声信号にRWCPデータベースのマイクアレイインパルス応答を畳み込んで作成, 2チャンネルで2音源の混合信号	JR2 impulse response 残響時間: 470 ms
テスト用音声信号	SISEC2011の音声信号にRWCPデータベースのマイクアレイインパルス応答を畳み込んで作成, 2チャンネルで2音源の混合信号	
比較手法	ILRMA (基底数2, 3及び4) の計4手法	
FFT長	8192 点 (512 ms)	
窓関数長	2048 点 (128 ms, ハミング窓)	
主観評価値	平均SDR改善値	



• 実験結果



DNNパーミュテーションソルバがILRMAと比較して**8 dB以上のSDR改善量**を示す
多数決によって, DNNの推定ミスに頑健なパーミュテーション解決が可能