

1. 研究背景

多チャンネル音源分離

- 多チャンネル観測信号をもとに特定の信号を推定

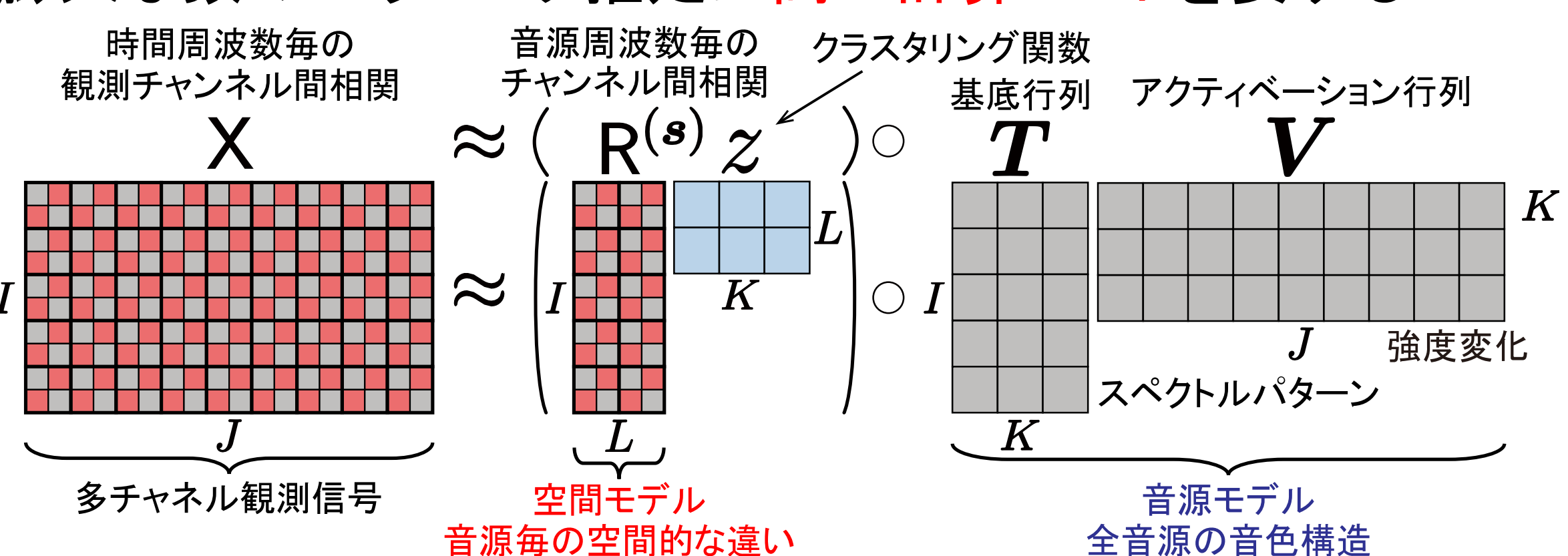


周波数領域での多チャンネル音源分離

- 短時間フーリエ変換して得られるスペクトログラムに対して周波数毎の分離フィルタを推定する手法が一般的

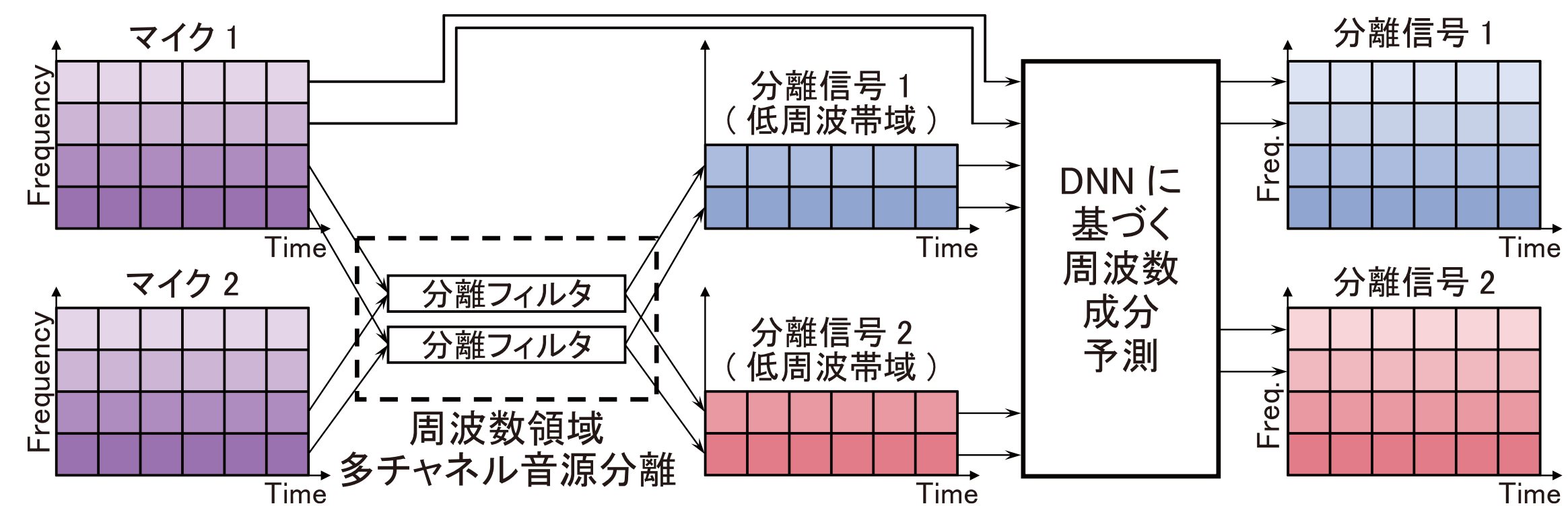
多チャンネル非負値行列因子分解(MNMF) [Sawada+, 2013]

- 事前情報なしで高品質な音源分離
- 膨大な数のパラメータ推定に高い計算コストを要する



本研究の目的

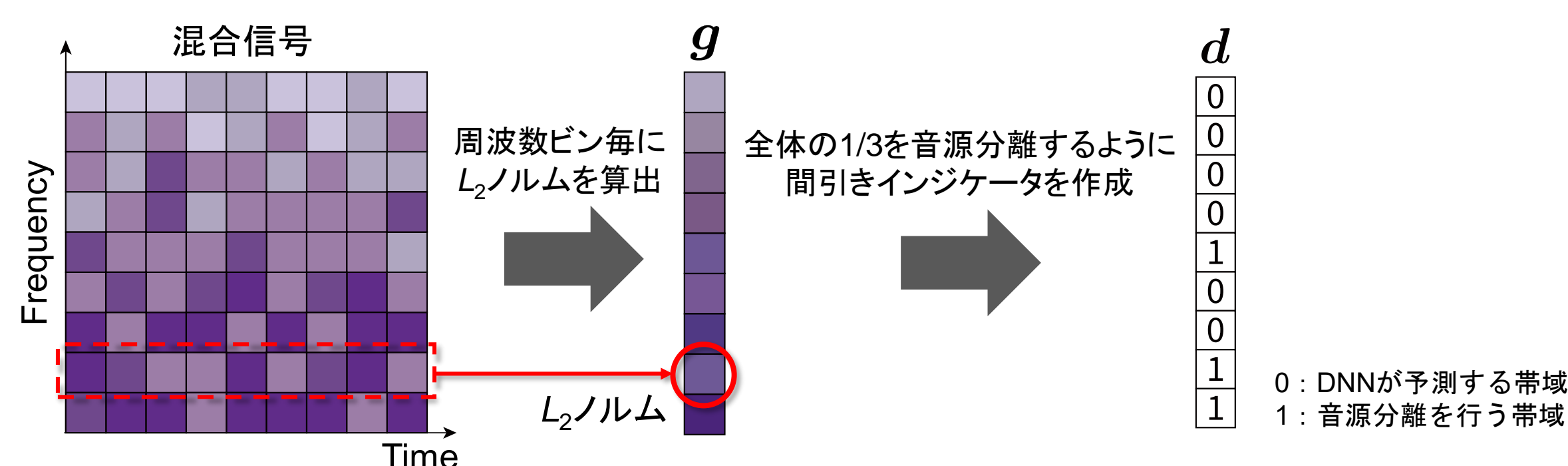
- 高品質な多チャンネル音源分離を低い計算コストで実現
- 周波数領域多チャンネル音源分離と深層学習(DNN)を組み合わせた音源分離フレームワークを提案 [Watanabe+, ASJ2021秋]
 - 特定の周波数帯域のみ先に分離, 他の帯域はDNNで予測・分離
 - 前段の混合信号を間引く処理が, 音源分離の品質に影響
- 本発表では, 実験を通してMNMFとの詳細な比較を行う



2. 提案手法

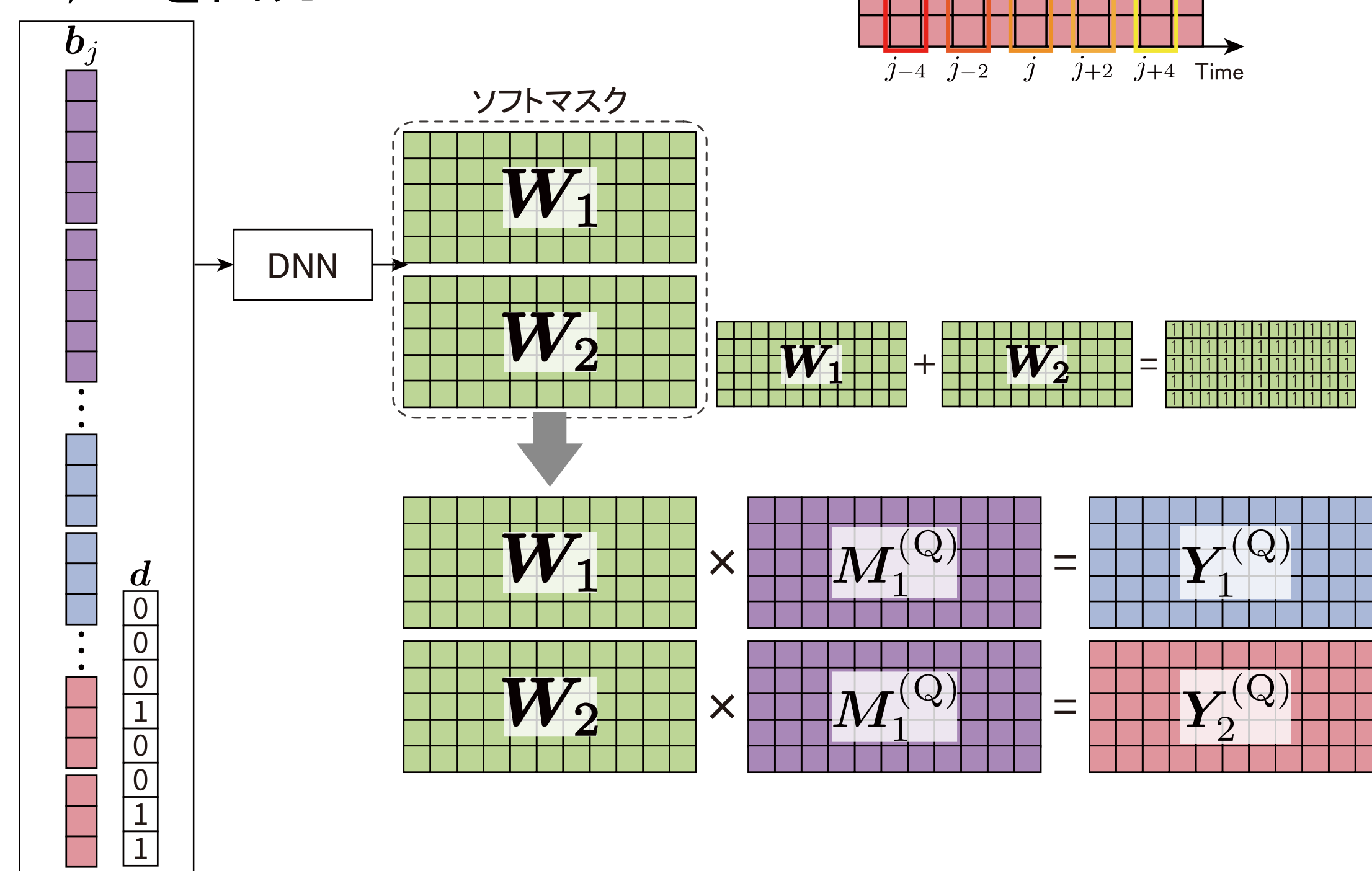
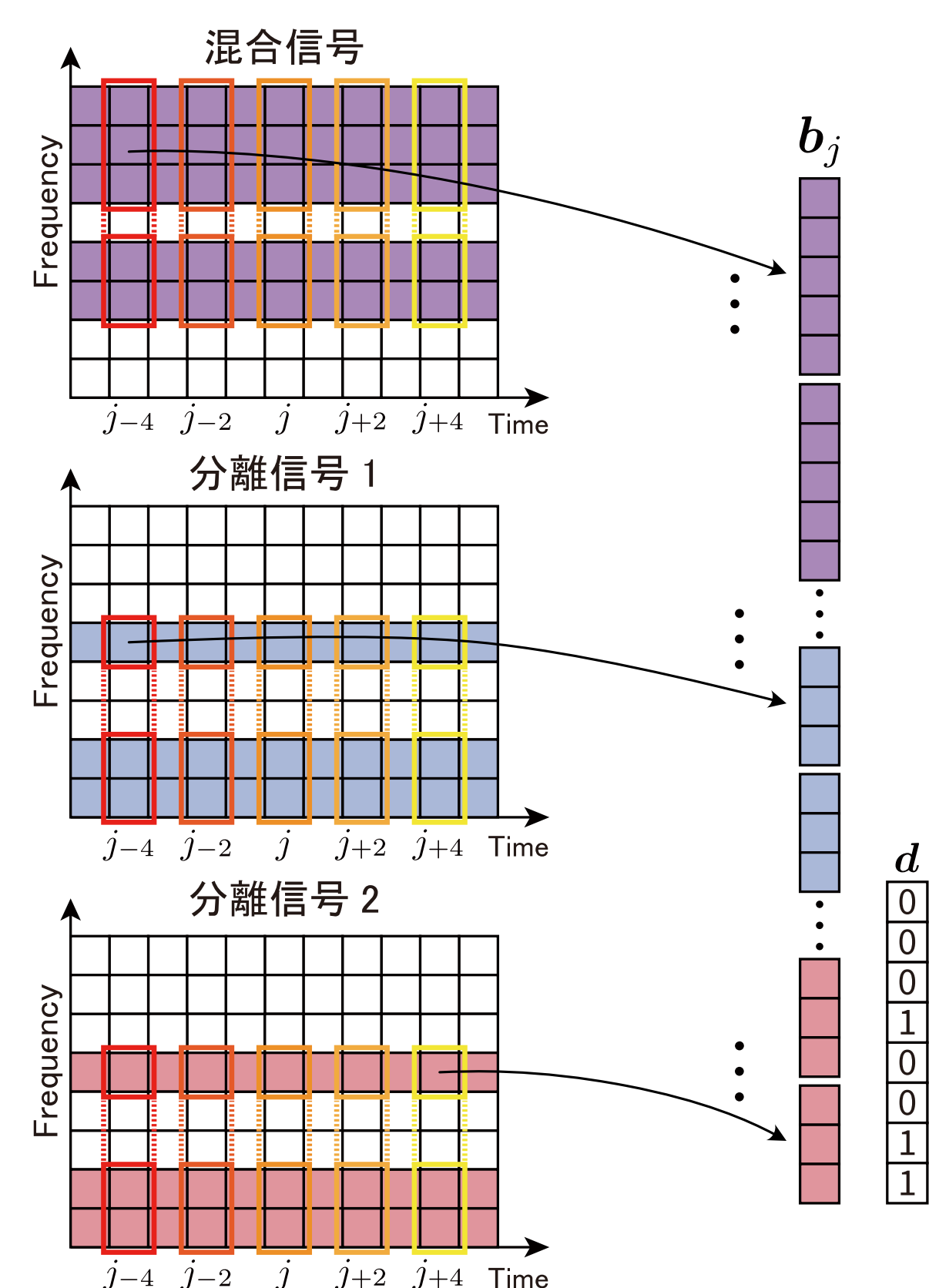
前段の混合信号を間引く処理

- 間引く帯域を示すインジケータにより前段の音源分離処理を行う
- 高周波帯域の間引き, 等間隔間引き, 不等間隔間引き, パワーに基づく間引きの4種類を提案 [Watanabe+, ASJ2021春] [Watanabe+, ASJ2021秋]
- 下図はパワーに基づく間引き手法でのインジケータ作成例



後段のDNNによる予測・分離

- 時間フレーム j に対して, 分離成分を予測
 - 各信号の振幅スペクトログラムの隣接時間フレームを連結しベクトル b_j を作成
 - ベクトル b_j と間引きインジケータ d を結合
- 結合したベクトルをDNNに入力
- DNNは $M^{(Q)}$ から, 音源分離を適用しなかった帯域 $Y_1^{(Q)}, Y_2^{(Q)}$ を得られるようなソフトマスク W_1, W_2 を出力



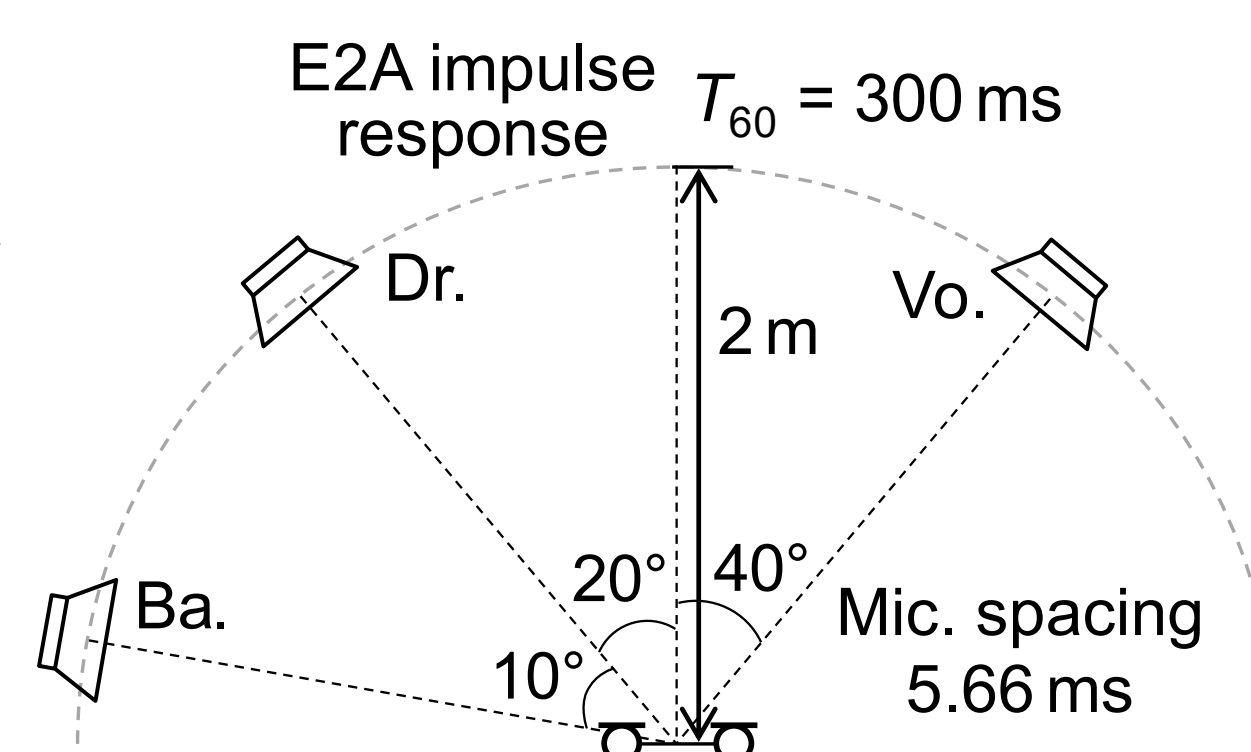
3. 実験

- 全周波帯域を分離するMNMF及び4つの間引き手法を用いた音源分離フレームワークとの比較を行う

Source-to-distortion-ratio (SDR) [Vincent+, 2016] の改善量と時間

実験条件

- 音源信号
 - SiSEC2018 MUSDBのベース(Ba.), ドラム(Dr.), 及びボーカル(Vo.) 10曲 ($F_s=16$ kHz)
- 混合条件
 - 2音源2マイクでの全組み合わせ及び3音源2マイクの4パターン
 - RWCP E2Aインパルス応答
 - 残響時間: 300 ms
- その他条件

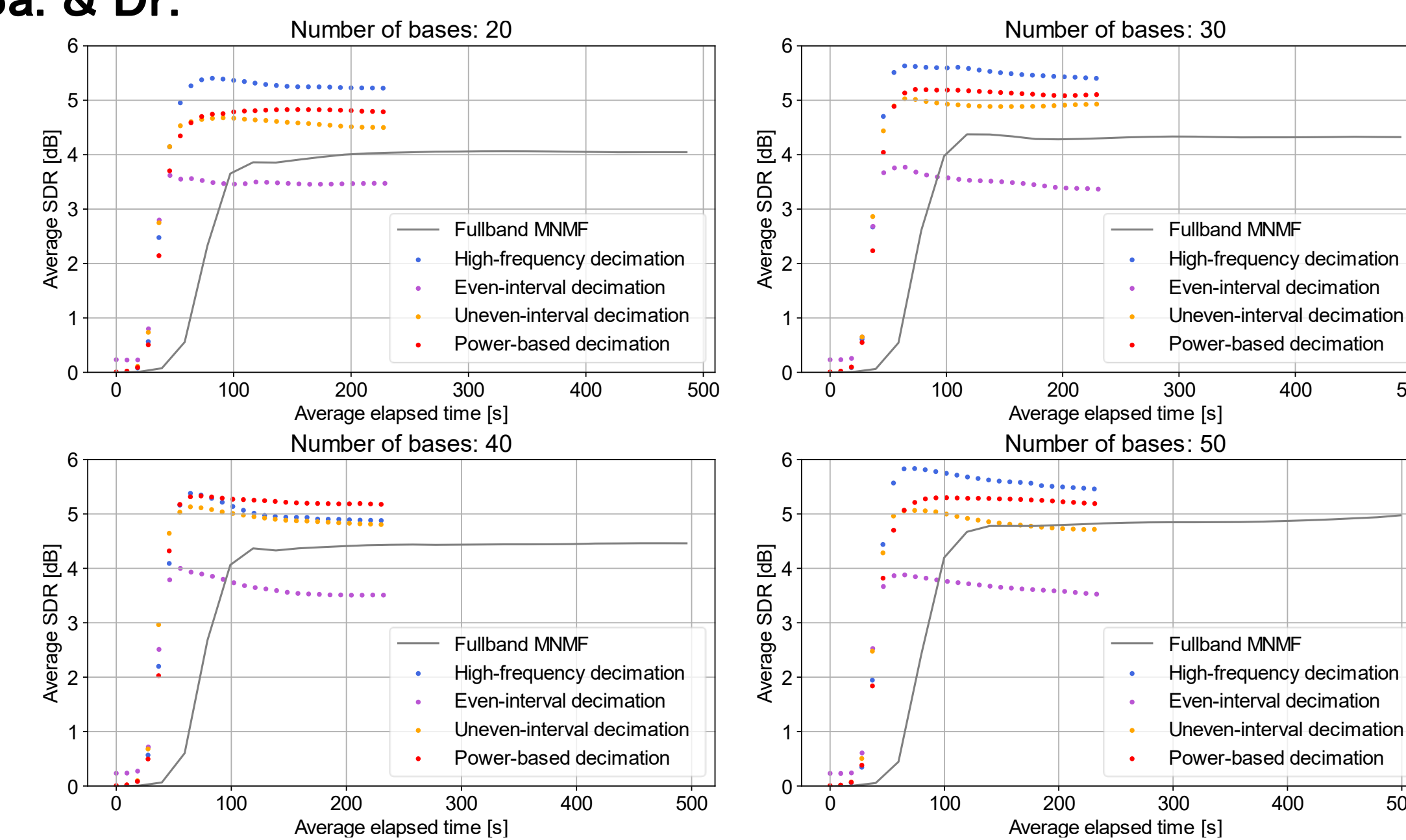


混合信号の間引き率	1/3
FFT窓長/シフト長	128 ms / 64 ms
スペクトログラムのビン数	1025
反復回数	500回
試行回数	異なる乱数シードで5回
基底数	20, 30, 40, 50

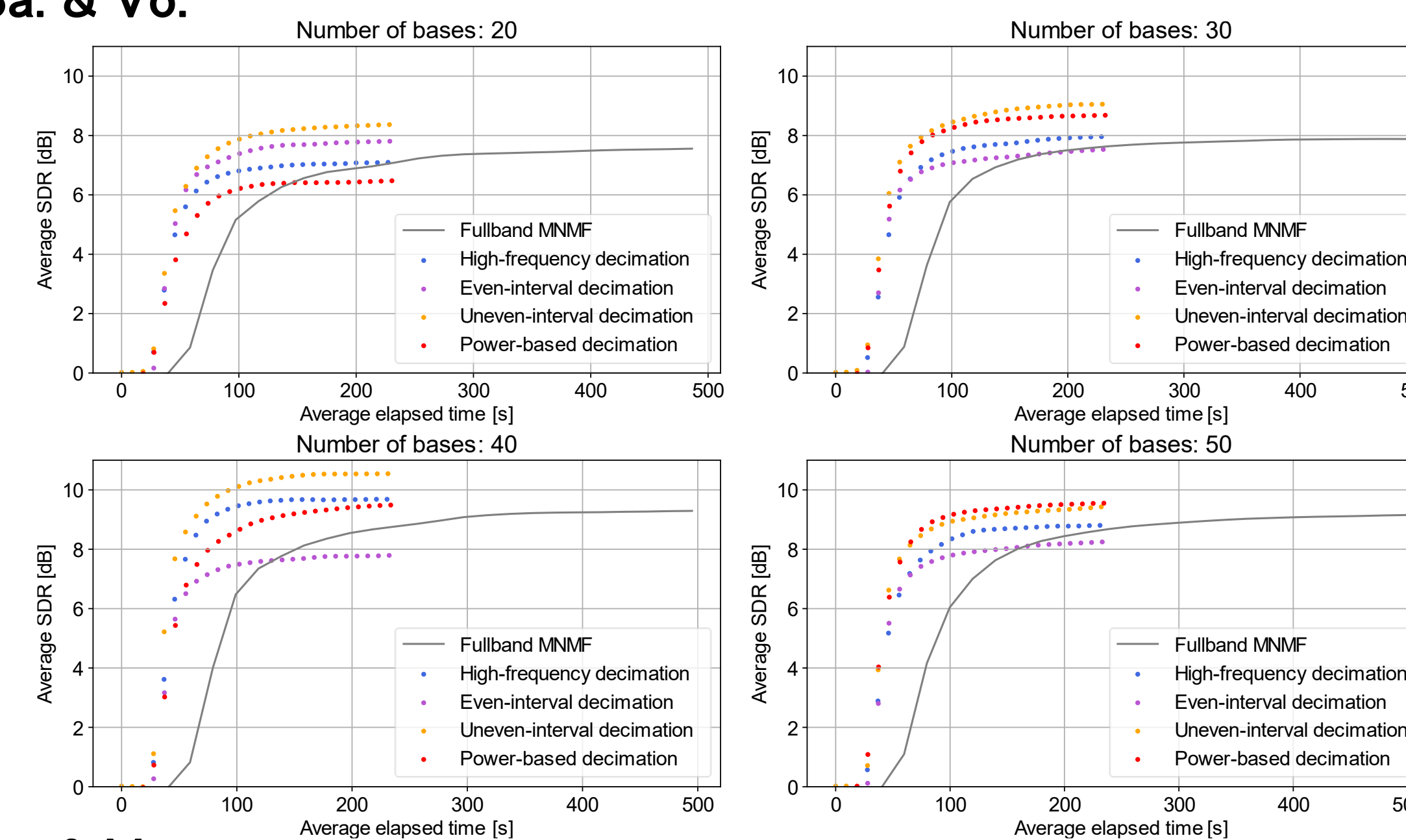
4. 実験結果

テストデータ10曲の平均SDR改善量の推移と処理時間の関係

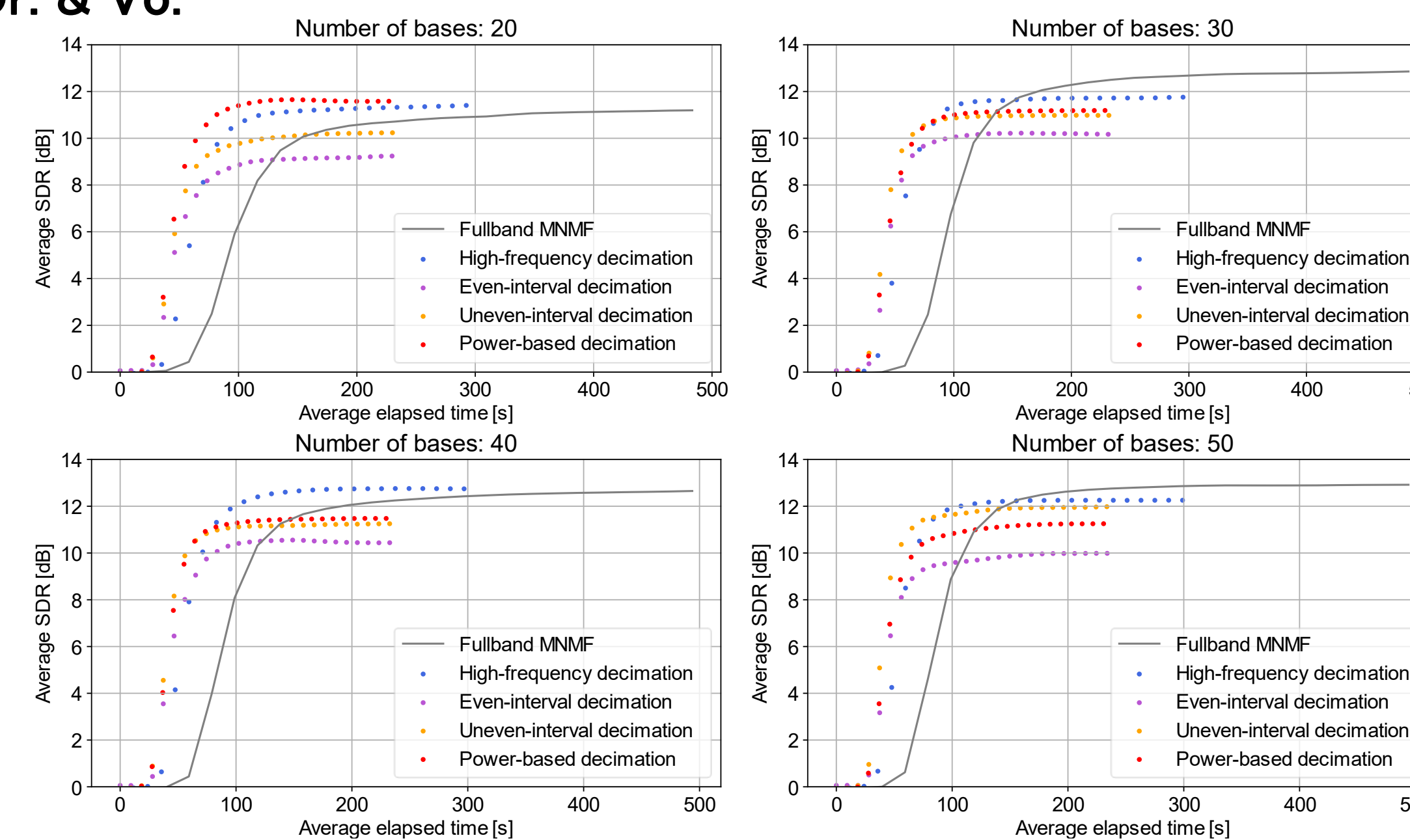
Ba. & Dr.



Ba. & Vo.



Dr. & Vo.



Ba. & Dr. & Vo.

