

©西川 剛樹† 荒木 章子‡ 牧野 昭二‡ 猿渡 洋†

(† 奈良先端大・情報, ‡ NTT コミュニケーション科学基礎研究所)

1 はじめに

複数の音源信号が混在されて観測された場合に、観測信号のみから音源信号を同定する技術を Blind Source Separation (BSS) と呼び、独立成分分析 (Independent Component Analysis: ICA) の観点から音源を分離する手法 [1] が多く検討されている。この ICA に基づく BSS においては、残響が少ない環境では十分な分離性能が得られている。しかし、残響環境下での音源分離に関しては分離性能が悪く、実環境下での音源分離問題に適用することは非常に困難である。

一般に残響除去を行うようなフィルタを構成する際、フィルタの分析フレーム長 (帯域分割数) を長くすると、残響除去性能が向上するということが知られている。しかしながら、BSS における逆混合行列は帯域分割数を過度に増やすと、分離性能は劣化してしまう [2]。

本研究では、この原因を、帯域分割した際の狭帯域信号間の独立性が低くなったためではないかと考えた。そこで、信号間の独立性を定量化し、帯域分割数と独立性の関係について調査を行った。

2 帯域分割型 ICA に基づく音源分離手法

L 個の音源信号が線形に混合されて K 個のマイクロホンで観測されたとする。時間 t 、周波数 f における観測信号を $X(f, t) = [X_1(f, t), \dots, X_K(f, t)]^T$ と表し、逆混合行列 $W(f)$ を用いて信号分離を行う。逆混合行列 $W(f)$ は、 L 個の出力 $Y(f, t) = [Y_1(f, t), \dots, Y_L(f, t)]^T$ が互いに独立となるように学習される。この処理は、以下のように示される。

$$Y(f, t) = W(f)X(f, t) \quad (1)$$

上式における逆混合行列を推定する方法としては Amari らによって提案された Natural gradient に基づく以下の更新式 [3] が用いられる。

$$W_{i+1}(f) = \eta \left[\text{diag} \left(\langle \Phi(Y(f, t)) Y^H(f, t) \rangle_t \right) - \langle \Phi(Y(f, t)) Y^H(f, t) \rangle_t \right] W_i(f) + W_i(f) \quad (2)$$

ただし $\langle \cdot \rangle_t$ は、入力信号をフレーム分析した際の全フレームでの平均である。また、 i は反復における i 回目の値を示しており、 η はステップサイズパラメータである。ここで、非線形ベクトル関数 $\Phi(\cdot)$ は、以下のように定義される。

$$\Phi(Y(f, t)) \equiv [\Phi(Y_1(f, t)), \dots, \Phi(Y_L(f, t))]^T \quad (3)$$

$$\Phi(Y_i(f, t)) \equiv [1 + \exp(-Y_i^{(R)}(f, t))]^{-1} + j \cdot [1 + \exp(-Y_i^{(I)}(f, t))]^{-1} \quad (4)$$

ただし、 $Y_i^{(R)}(f, t)$ と $Y_i^{(I)}(f, t)$ はそれぞれ $Y_i(f, t)$ の実数部分と虚数部分である。

3 帯域分割数と狭帯域音源間の独立性との関係

3.1 独立性の定量化

本研究では、帯域分割数を過度に増やすと分離性能が劣化する原因を、帯域分割した際の狭帯域信号間の独立性が

低下したためと考える。そこで、独立性を定量化し帯域分割数と独立性の関係を調べる。

独立性の評価には式 (2) における $[\cdot]$ 項を用い、改めて次式のように定義する。

$$V(f) \equiv \text{diag} \left(\langle \Phi(Y(f, t)) Y^H(f, t) \rangle_t \right) - \langle \Phi(Y(f, t)) Y^H(f, t) \rangle_t \quad (5)$$

$$= \begin{bmatrix} 0 & v_{12}(f) & \cdots & v_{1L}(f) \\ v_{21}(f) & 0 & & v_{2L}(f) \\ \vdots & & \ddots & \\ v_{L1}(f) & \cdots & v_{LL-1}(f) & 0 \end{bmatrix}$$

$$v_{lk}(f) \equiv -\langle \Phi(Y_l(f, t)) Y_k(f, t) \rangle_t \quad (6)$$

ここで、 $V(f)$ はある周波数 f における ICA の更新の増加分を表している。ICA は分離信号が独立になるように逆混合行列を更新するが、本更新を行うにつれて、行列 $V(f)$ のノルム (Frobenius norm) は小さくなっていく。そして、独立になると $V(f)$ は 0 行列 (ノルム 0) になる。つまり、 $V(f)$ のノルムが 0 からどの程度離れているかが、独立性を表しているといえる。よって $V(f)$ のノルムを独立性の評価値とする。

具体的な独立性の評価関数を I とし、次式のように定義する。

$$I = \langle \|V(f)\| \rangle_f = \left\langle \sqrt{\sum_{l \neq k} |v_{lk}(f)|^2} \right\rangle_f \quad (7)$$

ここで、狭帯域信号 $Y(f, t)$ は時間軸に対してパワー正規化してあるものとする。独立性の評価方法としては、式 (7) の I の値が大きい場合は信号間の独立性が低い、小さい場合は独立性が高いとみなす。

3.2 独立性の評価実験条件

本節では、音源信号として、ASJ 研究用連続音声コーパスより選択した男性 2 名による 2 短文を使用した。各信号の標準化周波数は 8 kHz であり、分析条件として、フレームシフト 16 (2 msec)、分析窓関数はハミング窓、本評価に用いる音源信号長は 3 秒である。

帯域分割数を 32 から 2048 (フレーム長は 4 から 192 msec に相当) の範囲で変化させ、式 (7) を用いて分割数と独立性の関係を調べた。ただし、等価な比較を行うためにフレーム数はどの帯域分割数の場合も同数とした。

3.3 独立性の評価結果

32 分割、2048 分割の場合の左右の音源における狭帯域信号 (1 kHz) の実部をそれぞれ図 1, 2 に示し、帯域分割数と独立性の関係を図 3 に示す。

図 3 より、帯域分割数を増やすと独立性が低下していくことが分かる。これは分割数 32 の場合の図 1 と分割数 2048 の場合の図 2 の狭帯域信号を比較しても分かる。つまり、分割数が少ない場合は狭帯域信号間の相関が低く、分割数が多い場合は信号間の相関が強いということが視覚的に確認できる。この結果より、音源分離実験において極端に帯域分割数を増やしても分離性能が向上しないのは、分割数を増やすぎると、狭帯域信号間の独立性が低くなってしまっているからだといえる。

3.4 円状畳み込みの影響

図 3 より、帯域分割数が少ない場合は独立性は高いことが分かる。しかし、分割数と分離性能の結果では、分割数が少なすぎると分離性能は逆に低くなることが分かっている

*Optimization on the Number of Subbands in Blind Source Separation with Subband ICA, by T. Nishikawa†, S. Araki‡, S. Makino‡, and H. Saruwatari† († Nara Institute of Science and Technology, ‡ NTT Communication Science Laboratories).

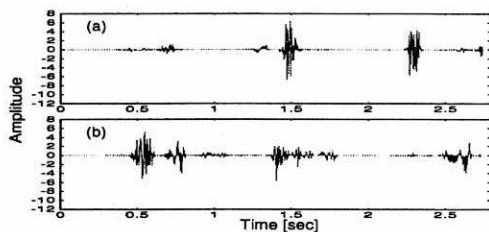


図 1: (a) 32 分割の場合の音源 (左) の狭帯域信号 (実部), (b) 32 分割の場合の音源 (右) の狭帯域信号 (実部)

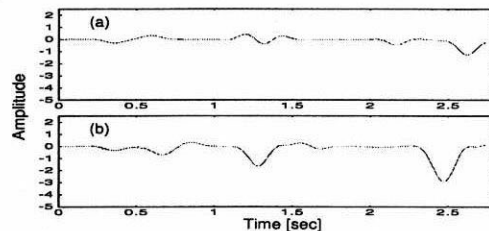


図 2: (a) 2048 分割の場合の音源 (左) の狭帯域信号 (実部), (b) 2048 分割の場合の音源 (右) の狭帯域信号 (実部)

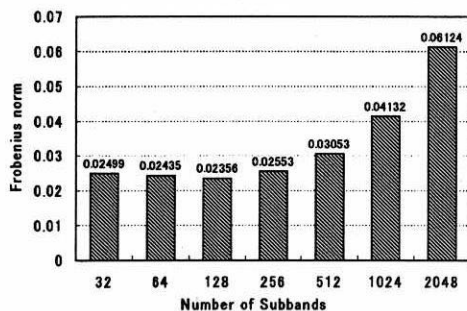


図 3: 帯域分割数と独立性の関係

[2]. この原因としては、円状畳み込みの影響が考えられる。そこで、円状畳み込みの影響を調査するために死角制御型ビームフォーミングを用いる。死角制御型ビームフォーミングとは、干渉音の方位に死角を向け、目的方位の音源を精度良く推定する技術である。特に無残響の場合には、音源方位が既知であれば、死角制御型ビームフォーミングを用いた音源分離が分離性能の上限を与える。よって、この場合は円状畳み込みの影響をみることができる。

円状畳み込みの影響のみを調べるために、音源の正確な方位は事前に分かっているものとして、この方位に死角を向けたビームフォーミング (理想死角制御型ビームフォーミング) を用いて音源分離を行う。分析条件としては、フレームシフト 16 サンプル (2 msec)、分析窓関数はハミング窓とした。3.2 節で用いた無残響下での観測信号に対して、帯域分割数を 32 から 2048 まで変化させ分離性能を調べた。

3.5 理想死角制御型ビームフォーミングによる音源分離実験結果

本実験では、分離性能の客観評価尺度として Noise Reduction Rate (NRR; $\equiv$ 出力 SNR [dB] - 入力 SNR [dB]) を用いる。

理想死角制御型ビームフォーミングを用いて音源分離を行った場合の帯域分割数と分離性能の関係を図 4 に示す。

本図より、帯域分割数が少ないほど分離性能が低いことが分かる。これは、分割数が少なすぎると、円状畳み込みの影響が大きくなり、高精度な分離が困難になるからだと推測できる。よって、帯域分割数が少ない場合は独立性は高いが、円状畳み込みの影響により分離性能は低くなるという。

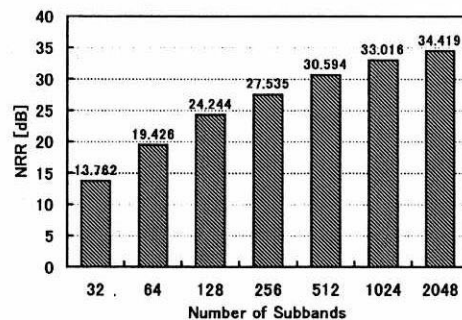


図 4: 理想死角制御型ビームフォーミングを用いた場合の帯域分割数と分離性能の関係 (無残響)

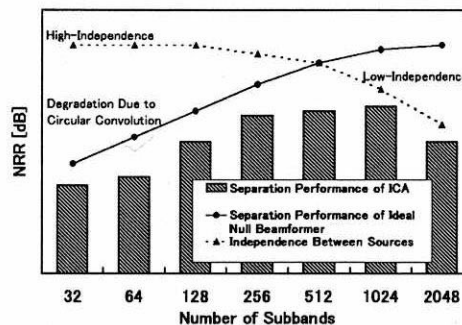


図 5: 帯域分割数と分離性能の関係を表す模式図。帯域分割数が少ない場合は円状畳み込み、帯域分割数が多い場合は狭帯域信号間の独立性に影響されることを表している。

3.6 周波数領域 ICA における最適な帯域分割数

周波数領域 ICA では独立性と円状畳み込みに関して以下の関係が存在する。

- (1) 帯域分割数を増やす場合、狭帯域信号間の独立性が低下する (図 3 参照)。よって、過度に帯域分割数を増やしても、分離性能は向上しない。
- (2) 帯域分割数が過度に少ない場合、円状畳み込みの影響により、また分離性能が劣化する (図 4 参照)。

以上 (1), (2) 間のトレードオフにより、分離性能が決まることが実験的に確認された。このトレードオフ関係を表した模式図を図 5 に示す。

なお、残響が存在する場合には、残響に対する逆フィルタを構成する意味において、帯域分割数は多いほうが好ましい。しかし、上記 (1) の性質により、ある一定以上の帯域分割数を用いることは必ずしも分離性能向上にはつながらない。

以上の結果より、残響環境下での音源分離については、周波数領域 ICA では最適な帯域分割数が存在する。

4 まとめ

本稿では、周波数領域 ICA における帯域分割数と狭帯域信号間の独立性の関係、円状畳み込みによる分離性能劣化について述べた。分割数を過度に増やすと分離性能が劣化する原因は、狭帯域信号間の独立性が低下してしまったためであることが独立性の評価実験結果より確認された。また、過度に分割数が少ないと独立性は高いが、円状畳み込みの影響により、分離性能は劣化することも確認された。以上の結果より、周波数領域 ICA では最適な帯域分割数が存在することが分かった。

参考文献

- [1] A. Bell et al.: *Neural Computation*, vol.7, pp.1129–1159, 1995.
- [2] 荒木章子 他: 音講論集, 3-7-4, Mar. 2001.
- [3] S. Amari et al.: *Advanced in Neural Information Processing System 8*, pp.753–763, MIT Press, Cambridge MA, 1996.