

非定常スペクトルサブトラクションによる音源分離後の残留雑音除去*

○向井 良 荒木 章子 澤田 宏 牧野 昭二

(日本電信電話株式会社, NTT コミュニケーション科学基礎研究所)

1 はじめに

独立成分分析 (ICA) を用いたブラインド音源分離 (BSS) の性能は残響の多い環境で著しく低下する。我々はこれまでに、ICA を用いた周波数領域 BSS の分離性能をインパルス応答を用いて解析し、妨害音の直接音部分はほぼ完全に除去できているが、残響部分が除去しきれず、これが性能低下の主な要因であることを確認している [1]。

そこで、本稿では ICA を用いて妨害音の直接音部分を除去し、残留している残響部分のスペクトルを推定して除去する手法を提案し、その性能評価結果を示す。

一般に、スペクトルサブトラクション (SS) [2] は、目的信号の無音区間を利用して定常な雑音のスペクトルを推定して雑音を除去する手法であるが、提案手法では、目的音と妨害音が同時に存在している場合に BSS によって分離を行い、得られた分離音を用いて残留雑音のスペクトルを推定するため、目的音の無音区間を必要とせず、また、非定常な妨害音も除去できるという利点がある。

2 アルゴリズム

2.1 周波数領域 BSS

時間領域において遅れや畳み込みの影響を受けながらの混合は、周波数領域での瞬時混合に変換できるため、実環境においては周波数領域での BSS が有効である。以下では信号を周波数領域で考える。音源信号 $\mathbf{s}(\omega, t) = [s_1, s_2, \dots, s_N]^T$ 、観測信号 $\mathbf{x}(\omega, t) = [x_1, x_2, \dots, x_M]^T$ 、分離信号 $\mathbf{y}(\omega, t) = [y_1, y_2, \dots, y_N]^T$ 、混合行列 $\mathbf{H}(\omega)$ 、分離行列 $\mathbf{W}(\omega)$ とする。 $\mathbf{x}(\omega, t) = \mathbf{H}(\omega)\mathbf{s}(\omega, t)$ 、 $\mathbf{y}(\omega, t) = \mathbf{W}(\omega)\mathbf{x}(\omega, t) = \mathbf{W}(\omega)\mathbf{H}(\omega)\mathbf{s}(\omega, t)$ である。

ICA の学習アルゴリズムとして、

$$\mathbf{W}_{i+1} = \mathbf{W}_i + \mu[\mathbf{I} - \langle \Phi(\mathbf{y})\mathbf{y}^H \rangle]\mathbf{W}_i \quad (1)$$

を用いる。 $\mathbf{I}$ は単位行列、 μ はステップサイズ、 $\langle \cdot \rangle$ は時間平均である。非線形関数 $\Phi(\mathbf{y})$ は、対象が複素信号であることを考慮して、

$$\Phi(\mathbf{y}) = \tanh(g \cdot \text{abs}(\mathbf{y}))e^{j \arg(\mathbf{y})} \quad (2)$$

を用いる [3]。 g はゲインパラメータである。

2.2 非定常スペクトルサブトラクション

$\mathbf{G} = \mathbf{WH}$ とすると、 $y_i = \sum_{j=1}^N g_{ij}s_j$ と書ける。 y_i を、目的信号 s_i に起因する出力であるストレート成分 $y_i^{(s)}$ と、妨害音 $s_j (j \neq i)$ に起因するクロス成分 $y_i^{(c)}$ に分けて考えると、 $y_i = y_i^{(s)} + y_i^{(c)}$ 、 $y_i^{(s)} = g_{ii}s_i$ 、 $y_i^{(c)} = \sum_{j \neq i} g_{ij}s_j$ である。

*Suppression of residual cross-talk component using non-stationary spectral subtraction, Ryo Mukai, Shoko Araki, Hiroshi Sawada and Shoji Makino (NTT Communication Science Laboratories, NTT Corporation)

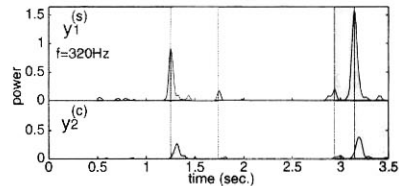


図 1: 狭帯域信号のストレートおよびクロス成分

ストレート成分 $y_i^{(s)}$ は残響や歪み等が付加されているため、 s_i の理想的な推定値ではないが、従来手法との比較のため、本稿では $y_i^{(s)}$ を抽出することを目指す。すなわち、除去されるべき残留雑音はクロス成分 $y_i^{(c)}$ とする。 $y_j (1 \leq j \leq N)$ を用いてクロス成分のスペクトル $|y_i^{(c)}|$ を推定し、 y_i から引くことで $\hat{y}_i^{(s)}$ を求める。

$$\hat{y}_i^{(s)}(\omega, t) = \begin{cases} (|y_i(\omega, t)|^\beta - |\hat{y}_i^{(c)}(\omega, t)|^\beta)^{1/\beta} \frac{y_i(\omega, t)}{|y_i(\omega, t)|} & (\text{if } |y_i(\omega, t)| > |\hat{y}_i^{(c)}(\omega, t)|) \\ 0 & (\text{otherwise}) \end{cases} \quad (3)$$

$\beta = 1$ の場合は振幅スペクトル、 $\beta = 2$ の場合はパワースペクトルによる処理になる。

2.3 残留雑音のモデル化

ストレート成分は音源信号の直接音および残響、クロス成分は音源信号の残響に起因しており、両者のスペクトルには関連がある。2入力2出力のICAによる分離信号の狭帯域信号のパワースペクトルの例を図1に示す。この例を見ると、ストレート成分がある遅延と係数をかけられてクロス成分に漏洩しているという見え方をしている。そこで、これを複数信号の場合に拡張し、各周波数ビンおよびチャンネルの組合せごとの遅延 $\tau_{ij}(\omega)$ および係数 $\alpha_{ij}(\omega)$ をパラメータとしたモデル化を行う。係数は本来はフィルタであるが、推定を簡単にするためにスカラー値とする。また、 $y_i^{(s)}$ は未知であるため、 y_i を $y_i^{(s)}$ の近似値として用いる。

$$|\hat{y}_i^{(c)}(\omega, t)|^\beta \approx \sum_{j \neq i} \alpha_{ij}(\omega) |y_j(\omega, t - \tau_{ij}(\omega))|^\beta \quad (4)$$

2.4 パラメータ推定方法

$|y_i^{(c)}(t)|$ と $|y_j^{(s)}(t - \tau_{ij})|$ との相関が大きくなるような遅延を求め、これを $\tau_{ij}(\omega)$ とする。 $y_i^{(c)}(t)$ と $y_j^{(s)}(t)$ は未知なので、 $y_i(t)$ と $y_j(t)$ で置き換えた (5) によって求める。これは、音源信号として互いに独立な音声を選定した場合、狭帯域信号 $|y_i^{(s)}|$ と $|y_j^{(s)}|$ が同時刻に値を持つことは少なく、ほとんどの時刻において

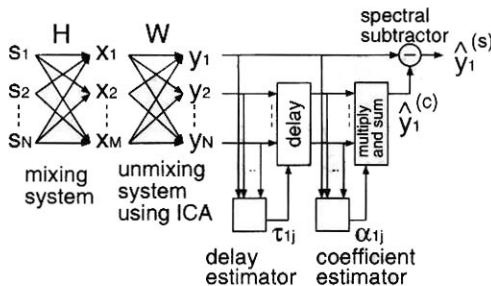


図 2: 提案手法のブロック図 ($i = 1$)

表 1: 実験条件

共通項目	sampling rate = 8kHz window = hanning
ICA 部	frame length=512 point (64ms) frame shift=128 point (16ms) $\mu = 0.1, g = 100.0$ number of iterations = 100
SS 部	frame length=1024 point (128ms) frame shift=64 point (8ms) $\beta = 2$

$|y_i(t)||y_j(t-\tau)|$ ($\tau > 0$) は $|y_i^{(c)}(t)||y_j^{(s)}(t-\tau)|$ で近似できるであろうという仮定による置き換えである。

$$\tau_{ij}(\omega) = \underset{0 < \tau \leq \tau_{max}}{\operatorname{argmax}} \sum |y_i(\omega, t)||y_j(\omega, t-\tau)| \quad (5)$$

ここで $\sum$ は、系が時不変であるとみなせる適当な期間、かつ $|y_i(t)| \leq |y_j(t-\tau)|$ が成り立っている時刻での総和とする。また、 τ_{max} は予想される残響時間の最大値から決まる定数である。

$\alpha_{ij}(\omega)$ は、 $|y_i^{(c)}(t)|^\beta$ と $|y_j^{(s)}(t-\tau_{ij})|^\beta$ の相関、および総和の比との積として求める。ここでもやはり $y_i^{(c)}(t)$ と $y_j^{(s)}(t)$ は未知なので、 $y_i(t)$ と $y_j(t)$ で置き換え、以下のように求める。

$$\alpha_{ij}(\omega) = \frac{\sum |y_i(\omega, t)|^\beta |y_j'(\omega, t)|^\beta}{\sqrt{\sum |y_i(\omega, t)|^{2\beta}} \sqrt{\sum |y_j'(\omega, t)|^{2\beta}}} \frac{\sum |y_i(\omega, t)|^\beta}{\sum |y_j'(\omega, t)|^\beta} \quad (6)$$

ここで、 $y_j'(\omega, t) = y_j(\omega, t - \tau_{ij}(\omega))$ 、 $\sum$ は $|y_i(\omega, t)|$ が $|y_j'(\omega, t)|$ より小さい時刻の集合

$$T_{ij}(\omega) = \{t \mid |y_i(\omega, t)| \leq |y_j'(\omega, t)|\} \quad (7)$$

に関する和 $\sum_{t \in T_{ij}(\omega)}$ である。これは、漏洩の方向（どちらがストレー成分か）をスペクトルの大きさで判断していることに相当する。

本アルゴリズムのブロック図 ($i = 1$ の部分のみ) を図 2 に示す。

3 実験

3.1 実験方法

音源数 2 の場合について実験を行った。実験条件を表 1 に示す。ASJ 研究用連続音声コーパス中の話

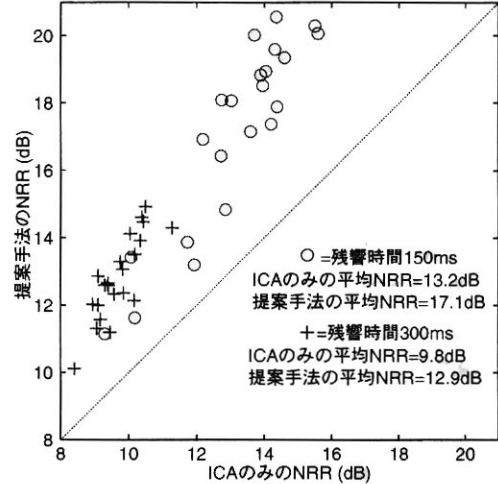


図 3: 実験結果

者 4 名 (男声 2, 女声 2), 2 通りの文の音声を用い、計 24 通りの組み合わせについて ICA のみによる分離性能と提案手法による性能を評価した。学習データ、評価データとも 6 秒である。

ストレー成分 $y_i^{(s)}$ をリファレンス信号として、出力 $\text{SNR}_i \equiv 10 \log(|y_i^{(s)}|^2 / |y_i^{(s)} - y_i^{(c)}|^2)$ (dB) から Noise Reduction Rate ($\text{NRR}_i \equiv \text{出力 SNR}_i - \text{入力 SNR}_i$) を求め、 NRR_1 と NRR_2 の平均値を評価尺度として用いた。

3.2 実験結果

実験結果を図 3 に示す。横軸は ICA の出力の NRR、縦軸は提案手法による NRR であり、音声の組合せごとに点をプロットした。すべての組合せにおいて性能が向上しており、NRR 改善量は残響時間 150ms の場合に 1.3~6.3dB (平均 3.9dB)、残響時間 300ms の場合に 1.7~4.4dB (平均 3.1dB) である。

また、SS においてしばしば問題となるミュージカルトーンノイズは発生しなかった。

4 まとめ

ICA を用いた BSS によって求めた分離音から、残留雑音であるクロス成分のスペクトルを推定して引き去ることによって分離性能を向上させる手法を提案し、残響下での混合音声を用いた実験によってその効果を確認した。

参考文献

- [1] R. Mukai, S. Araki, and S. Makino, "Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment," in *Proc. Eurospeech2001*, 2001.
- [2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, no. 2, pp. 113-120, April 1979.
- [3] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A polar-coordinate based activation function for frequency domain blind source separation," in *ICA2001*, 2001, (submitted).