

周波数領域ICAと時間遅れスペクトル減算による 残響下での実時間ブラインド音源分離*

○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

我々はこれまでに、独立成分分析 (ICA) を用いたブラインド音源分離 (BSS) の後処理として、分離音に含まれる残留雑音のスペクトルを推定して除去する手法である時間遅れスペクトルサブトラクション (Time-Delayed Spectral Subtraction, TDSS) を提案している [1] [2]。これまでの研究では、オフラインによる処理の性能評価を行い、その効果を確認しているが、音声認識システムのフロントエンドとして使用するような場合には、オンライン処理による実時間動作が必要である。そこで本稿では、TDSS のアルゴリズムをオンライン化し、市販の PC 上での実時間動作を確認したのでこれを報告する。

2 時間遅れスペクトルサブトラクション

ICA を用いた BSS の性能は、残響の多い環境で著しく低下する。ICA によって得られた分離系のインパルス応答を調べると、妨害音の残響部分がクロス成分 $y_i^{(c)}$ として残留していることがわかる (図 1) [3]。

次にこれを周波数領域で見ることにする。音源信号 $S(\omega, t) = [S_1, S_2, \dots, S_N]^T$, 観測信号 $X(\omega, t) = [X_1, X_2, \dots, X_M]^T$, 分離信号 $Y(\omega, t) = [Y_1, Y_2, \dots, Y_N]^T$, 混合行列 $H(\omega)$, 分離行列 $W(\omega)$ とする。 $X(\omega, t) = H(\omega)S(\omega, t)$, $Y(\omega, t) = W(\omega)X(\omega, t) = W(\omega)H(\omega)S(\omega, t)$ である。

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 $Y_i^{(c)}$ の和 $Y_i = Y_i^{(s)} + Y_i^{(c)}$ と考えると、 $Y_i^{(c)}$ が分離音に含まれる残留雑音である。

各周波数ビンの信号で $Y_i^{(c)}$ と $Y_j^{(s)}$ の関係を調べてみると、図 2 に見られるように、 $Y_j^{(s)}$ がある遅延 $\tau_{ij}(\omega)$ と係数 $\alpha_{ij}(\omega)$ をかけられて $Y_i^{(c)}$ に漏洩していることがわかる。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。DFT 長が長い場合には $Y_i^{(c)}$ と $Y_j^{(s)}$ はほぼ相似形になるため、スカラー値によるモデル化で十分近似できる。TDSS では、各周波数ビンごと、各チャンネルの組合せごとに遅延と漏洩係数を推定し、 $Y_i^{(c)}$ のスペクトルを推定して除去する。

2.1 アルゴリズムのオンライン化

オフラインアルゴリズム [1] [2] において「系が時不変と考えられる適当な期間」についての $Y_i(\omega, t)$ と

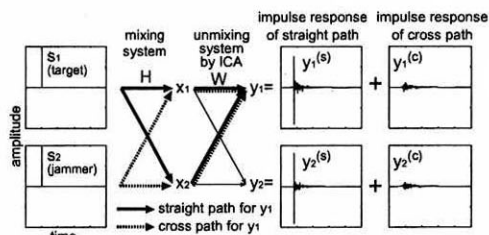


図 1: ストレートパスとクロスパスのインパルス応答

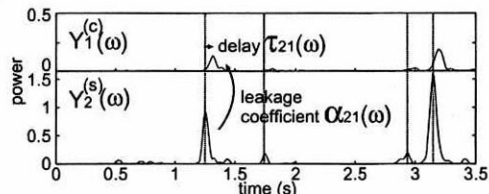


図 2: ストレート成分とクロス成分の関係

$Y_j(\omega, t)$ の相関やパワーの比を用いて計算していた部分を、忘却係数を用いた逐次更新に置き換えることでオンライン化する。

遅延 $\tau_{ij}(\omega)$ の推定

$\tau_{ij}(\omega)$ は $|Y_i^{(c)}(\omega, t)|$ と $|Y_j^{(s)}(\omega, t - \tau_{ij}(\omega))|$ との相関が大きくなるような正の遅延である。音源信号として互いに独立な音声を想定した場合、狭帯域信号 $|Y_i^{(s)}(\omega, t)|$ と $|Y_j^{(s)}(\omega, t)|$ が同時刻に値をもつことは少ない。そこで、 $|Y_i(\omega, t)| < |Y_j(\omega, t - \tau)|$ のときに $|Y_i^{(c)}(\omega, t)| |Y_j^{(s)}(\omega, t - \tau)| \approx |Y_i(\omega, t)| |Y_j(\omega, t - \tau)|$ という近似が成り立つと仮定し、間近の相互相関を測るための尺度 $C_{ij}(\omega, \tau)$ を以下の更新式によって求める。

$$\begin{aligned} \text{if } |Y_i(\omega, t)| < |Y_j(\omega, t - \tau)| \text{ then} \\ C_{ij}(\omega, \tau) \leftarrow \gamma_C C_{ij}(\omega, \tau) \\ + (1 - \gamma_C) |Y_i(\omega, t)| |Y_j(\omega, t - \tau)| \end{aligned} \quad (1)$$

ここで、 γ_C は忘却係数である。この尺度を最大にするような値を逐次計算して $\tau_{ij}(\omega)$ とする。

$$\tau_{ij}(\omega) = \underset{0 < \tau \leq \tau_{\max}}{\operatorname{argmax}} C_{ij}(\omega, \tau) \quad (2)$$

漏洩係数 $\alpha_{ij}(\omega)$ の推定

$Y_j(\omega, t)$ を $\tau_{ij}(\omega)$ だけ遅らせた信号を $Y_j'(\omega, t) = Y_j(\omega, t - \tau_{ij}(\omega))$ とする。ここで、 $|Y_i^{(c)}(\omega, t)| <$

*Real time blind source separation in reverberant environment using frequency domain ICA and time-delayed spectral subtraction, Ryo Mukai, Shoko Araki, Hiroshi Sawada and Shoji Makino (NTT Communication Science Laboratories, NTT Corporation)

表 1: 実験条件

共通項目	sampling rate = 8kHz window = hanning
ICA 部	frame length=512 point (64ms) frame shift=256 point (32ms) $\mu = 0.025$, $g = \infty$
TDSS 部	frame length=1024 point (128ms) frame shift=64 point (8ms) $\gamma_C = 0.95$, $\gamma_\alpha = 0.95$, $\beta = 1$

$|Y_j'(\omega, t)|$ のときに $|Y_i^{(c)}(\omega, t)|/|Y_j^{(s)}(\omega, t - \tau_{ij}(\omega))| \approx |Y_i(\omega, t)|/|Y_j'(\omega, t)|$ という近似が成り立つと仮定し、以下の更新式によって $\alpha_{ij}(\omega)$ を逐次求める。

$$\begin{aligned} & \text{if } |Y_i(\omega, t)| < |Y_j'(\omega, t)| \text{ then} \\ & \alpha_{ij}(\omega) \leftarrow \gamma_\alpha \alpha_{ij}(\omega) \\ & \quad + (1 - \gamma_\alpha) \left(\frac{|Y_i(\omega, t)|}{|Y_j'(\omega, t)|} \right)^\beta \end{aligned} \quad (3)$$

γ_α は忘却係数である。 $\beta = 1$ 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。

上記で求めた $\tau_{ij}(\omega)$, $\alpha_{ij}(\omega)$ を用いてクロス成分のスペクトルを、

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

のように推定し、 Y_i から引くことでストレート成分のスペクトルを求める。

$$\hat{Y}_i^{(s)}(\omega, t) = \begin{cases} (|Y_i(\omega, t)|^\beta - |\hat{Y}_i^{(c)}(\omega, t)|^\beta)^{1/\beta} \frac{Y_i(\omega, t)}{|Y_i(\omega, t)|} & \text{(if } |Y_i(\omega, t)| > |\hat{Y}_i^{(c)}(\omega, t)|) \\ 0 & \text{(otherwise)} \end{cases} \quad (5)$$

周波数領域 BSS

ICA による周波数領域 BSS の部分は以下の更新式を用いた [4]。

$$\mathbf{W}_{i+1} = \mathbf{W}_i + \mu [\mathbf{I} - \Phi(\mathbf{Y})\mathbf{Y}^H] \mathbf{W}_i \quad (6)$$

$\mathbf{I}$ は単位行列、 μ はステップサイズある。 μ の値をオフラインアルゴリズムのときよりも小さく設定することにより、パッチ処理における時間平均に相当する平滑化が行われる。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、

$$\Phi(\mathbf{Y}) = \tanh(g \cdot \text{abs}(\mathbf{Y})) e^{j \arg(\mathbf{Y})} \quad (7)$$

3 実験

2つの静止音源からの混合音をリアルタイムに分離する実験を行った。ハードウェアは市販の PC (CPU: AMD Athlon XP 1800+, クロック: 1533MHz, メモリ: 1.5GB, OS: Linux) を使用した。実験条件を表 1 に示す。

実験結果のスペクトログラムの一部を拡大したものを図 3 に示す。ICA のみによる出力 Y_1 中には妨害音 S_2 に起因するクロス成分が残留しているが、TDSS の出力 $\hat{Y}_1^{(s)}$ ではそれが抑圧され、目的音 S_1 に近い信号が得られている。

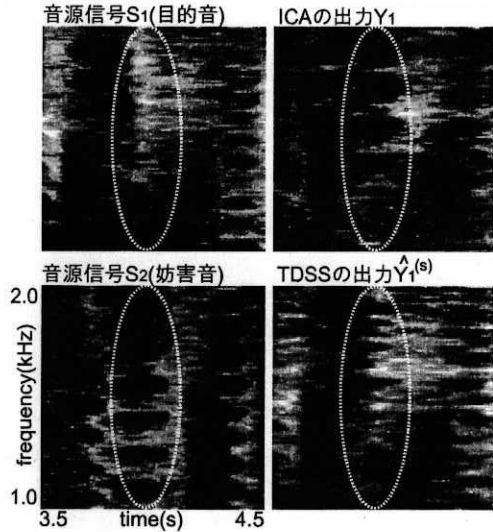


図 3: 実験結果 (残響時間 150ms)

4 まとめ

周波数領域 ICA と時間遅れスペクトルサブトラクション (TDSS) を用いた音源分離システムを実時間動作させ、残響下における混合音声の分離を行った。本手法は、ICA によって得られた分離音中に残留する雑音が、主に妨害音の残響であるという点に着目し、分離信号自身を用いて残留雑音成分を推定して除去するものである。これにより、残響に弱いという ICA の欠点を補うことができる。また、一般的なスペクトルサブトラクションと違い、他チャンネルの分離信号を用いることで非定常な雑音スペクトルを精密に推定できるため、引き過ぎによるミュージカルトーンノイズは発生しにくい。

参考文献

- [1] 向井 良, 荒木 章子, 澤田 宏, 牧野 昭二, “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,” in 音講論集 2-6-14, Oct. 2001, pp. 615-616.
- [2] R. Mukai, S. Araki, H. Sawada, and S. Makino, “Removal of residual cross-talk components in blind source separation using time-delayed spectral subtraction,” in *Proc. of ICASSP2002*, 2002, accepted.
- [3] R. Mukai, S. Araki, and S. Makino, “Separation and dereverberation performance of frequency domain blind source separation,” in *Proc. of Intl. Conf. on Independent Component Analysis and Blind Signal Separation (ICA2001)*, 2001, pp. 230-235.
- [4] S. Haykin, Ed., *Unsupervised adaptive filtering*, John Wiley & Sons, 2000.
- [5] H. Sawada, R. Mukai, S. Araki, and S. Makino, “Polar coordinate based nonlinear function for frequency-domain blind source separation,” in *Proc. of ICASSP2002*, 2002, accepted.