

集中マイクロホンの指向特性を利用したブラインド音源分離

松本 光春[†] 橋本 周司[†]

[†] 早稲田大学大学院理工学研究科物理学及応用物理学専攻

1 はじめに

筆者らは集中マイクロホンを用いた小型マイクロホンアレイを提案している [1]。本論文では、各マイクロホンの指向特性、及び、音響スペクトルが未知の状況でのブラインド音源分離アルゴリズムを提案する。また、楽器音や音声に対して行ったシミュレーション実験の結果を述べる。

2 問題設定

N 個の音源を同位置に置かれた M 個のマイクロホンで受信する問題を考える。 $\mathbf{X}(\omega, t) = [X_i(\omega, t)] (i = 1, 2, \dots, M)$ と $\mathbf{S}(\omega, t) = [S_j(\omega, t)] (j = 1, 2, \dots, N)$ を、それぞれ、受信信号ベクトル、及び、マイクロホン位置での音源信号ベクトルとする。このとき、受信信号は以下のように書ける。

$$\mathbf{X}(\omega, t) = \mathbf{M}(\omega)\mathbf{S}(\omega, t) \quad (1)$$

ただし、 ω は周波数を表し、 t は短時間フーリエ変換における時間依存を示す。 $\mathbf{M}(\omega) = [M_{ij}(\omega)]$ は混合行列であり、集中マイクロホンを用いる場合、 $\mathbf{M}(\omega)$ の要素 $M_{ij}(\omega)$ はマイクロホンの伝達関数となる。このとき、分離過程は以下のように定式化される。

$$\mathbf{Y}(\omega, t) = \mathbf{W}(\omega)\mathbf{X}(\omega, t) \quad (2)$$

ここで、 $\mathbf{Y}(\omega, t)$ 、 $\mathbf{W}(\omega)$ はそれぞれ分離信号ベクトル、及び、分離行列を示す。本論文では、2入力2出力 ($N=M=2$) の状況を考え、 $\mathbf{M}(\omega)$ 、 $\mathbf{S}(\omega, t)$ が未知の状態でも音源信号の振幅成分を復元する問題を取り扱う。

3 分離手法

3.1 音源に対する仮定

本手法では、音源に対し、以下の3つの仮定をする。仮定1：周波数軸上でスペクトルが重ならない点を持つ。

これは、以下を満たす周波数 ω_α 、 ω_β が観測時間中に存在することに対応する。

$$\begin{cases} |S_1(\omega_\alpha, t)| \neq 0 \\ |S_2(\omega_\alpha, t)| = 0 \end{cases} \quad (3)$$

A blind decomposition utilizing the directivity of aggregated microphones

[†] Mitsuharu MATSUMOTO(matsu@shalab.phys.waseda.ac.jp)

[†] Shuji HASHIMOTO(shuji@waseda.jp)

Major in Pure and Applied Physics, Graduate school of Science and Engineering, Waseda University (†)

$$\begin{cases} |S_1(\omega_\beta, t)| = 0 \\ |S_2(\omega_\beta, t)| \neq 0 \end{cases} \quad (4)$$

ただし ω_α 、 ω_β をあらかじめ知っている必要はない。仮定2：2つの指向性マイクロホンの指向特性の比が、2つの音源方向に対して異なっている。

$d_{ij} = |M_{ij}(\omega)|$ を、 i 番目のマイクロホンの j 番目の音に対する利得とする。2つのマイクロホンの利得の比 D_j を以下のように定義する。

$$D_j = \frac{d_{2j}}{d_{1j}} \quad (5)$$

仮定2は以下を満たされることに対応する。

$$D_1 \neq D_2 \quad (6)$$

説明のため、ここでは $D_1 > D_2$ とする。

仮定3：混合音の振幅が混合前の音源の振幅の和の形で記述できる。

これは、 $|X_1(\omega, t)|$ 、 $|X_2(\omega, t)|$ について以下が成立することに対応する。

$$|X_1(\omega, t)|^2 = d_{11}^2 |S_1(\omega, t)|^2 + d_{12}^2 |S_2(\omega, t)|^2 \quad (7)$$

$$|X_2(\omega, t)|^2 = d_{21}^2 |S_1(\omega, t)|^2 + d_{22}^2 |S_2(\omega, t)|^2 \quad (8)$$

簡単のため、式(7)、(8)を以下のように変形する。

$$|X_1(\omega, t)|^2 = |S'_1(\omega, t)|^2 + |S'_2(\omega, t)|^2 \quad (9)$$

$$|X_2(\omega, t)|^2 = D_1^2 |S'_1(\omega, t)|^2 + D_2^2 |S'_2(\omega, t)|^2 \quad (10)$$

ただし、

$$S'_1(\omega, t) = d_{11} S_1(\omega, t) \quad (11)$$

$$S'_2(\omega, t) = d_{12} S_2(\omega, t) \quad (12)$$

$S'_j(\omega, t)$ は、 $S_j(\omega, t)$ は定数倍した関数である。そのため、音源分離問題を $X_1(\omega, t)$ 、 $X_2(\omega, t)$ から、 $S'_1(\omega, t)$ 、 $S'_2(\omega, t)$ を復元する問題としても一般性を失わない。

3.2 指向特性 D_1 、 D_2 の特定

受信信号のパワー比 $\Delta A(\omega, t)$ を以下で定義する。

$$\Delta A(\omega, t) = \frac{|X_2(\omega, t)|}{|X_1(\omega, t)|} \quad (13)$$

$\Delta A(\omega, t)$ は以下のように書ける。

$$\begin{aligned}\Delta A(\omega, t)^2 &= \frac{D_1^2 |S'_1(\omega, t)|^2 + D_2^2 |S'_2(\omega, t)|^2}{|S'_1(\omega, t)|^2 + |S'_2(\omega, t)|^2} \\ &= D_1^2 + \frac{(D_2^2 - D_1^2) |S'_2(\omega, t)|^2}{|S'_1(\omega, t)|^2 + |S'_2(\omega, t)|^2} \\ &= D_2^2 + \frac{(D_1^2 - D_2^2) |S'_1(\omega, t)|^2}{|S'_1(\omega, t)|^2 + |S'_2(\omega, t)|^2}\end{aligned}\quad (14)$$

$|S'_1(\omega, t)|$ 、 $|S'_2(\omega, t)|$ は非負なので、 $\Delta A(\omega, t)$ は以下のように書ける。

$$\Delta A(\omega, t) \begin{cases} < D_1 & (\omega \neq \omega_\alpha) \\ = D_1 & (\omega = \omega_\alpha) \end{cases}\quad (15)$$

$$\Delta A(\omega, t) \begin{cases} > D_2 & (\omega \neq \omega_\beta) \\ = D_2 & (\omega = \omega_\beta) \end{cases}\quad (16)$$

D_1 、 D_2 の予測値 \overline{D}_1 、 \overline{D}_2 は以下のように求められる。

$$\overline{D}_1 = \min[\Delta A(\omega, t)] = \min\left[\frac{|X_2(\omega, t)|}{|X_1(\omega, t)|}\right]\quad (17)$$

$$\overline{D}_2 = \max[\Delta A(\omega, t)] = \max\left[\frac{|X_2(\omega, t)|}{|X_1(\omega, t)|}\right]\quad (18)$$

3.3 音源分離

音源 $S'_i(t)$ は以下のようにして復元される。

$$|S'_1(\omega, t)| = \frac{|X_2(\omega, t)| - \overline{D}_2 |X_1(\omega, t)|}{\overline{D}_1 - \overline{D}_2}\quad (19)$$

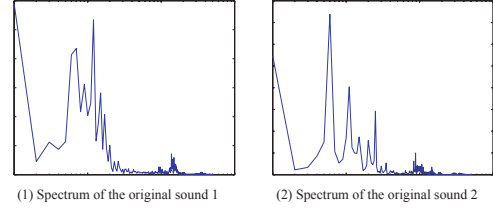
$$|S'_2(\omega, t)| = \frac{|X_2(\omega, t)| - \overline{D}_1 |X_1(\omega, t)|}{\overline{D}_2 - \overline{D}_1}\quad (20)$$

4 実験

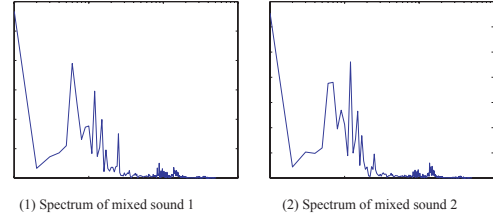
記述したアルゴリズムの有効性を確かめるため、音源分離に関するシミュレーション実験を行った。実験では反響のない自由音場を設定し、マイクロホンとしては、単一指向性マイクロホンと無指向性マイクロホンの組を仮定した。一般に音圧一次傾度形単一指向性を持つマイクロホンの利得 $d(\theta)$ は以下のリマソン曲線で近似できることが知られる。

$$d(\theta) = a + b \cos(\theta) \quad 0 < b < a \quad (21)$$

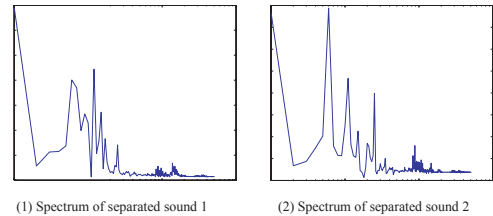
ただし、 θ は入射角を示す。ここでは、岡田らの研究に基づき、 a 、 b をそれぞれ 0.6、0.4 に設定した [2]。また、無指向性マイクロホンの利得を 0.8 に設定した。音源方向は、一音源を指向性マイクロホンの軸方向に、他の一音源を軸方向から 45 度方向に設定した。ここでは 2 音に男性音を用いた場合の結果を示す。音源には同一人物から発声された日本語 ”い” と ”え” の音を用いた。図 1(a)、図 1(b)、図 1(c) はそれぞれ原信号、混合信号、分離音の時間軸上での波形を示す。 D_1 、 D_2 の値 1.25、0.89 に対し、予測された \overline{D}_1 、 \overline{D}_2 は、それぞれ 1.1095、0.8262 であった。



(a) Spectra of the original sounds.



(b) Spectra of the mixed sounds



(c) Spectra of the separated sounds by the proposed blind decomposition algorithm. $D_1 = 1.25$, $D_2 = 0.89$, $\overline{D}_1 = 1.1095$, $\overline{D}_2 = 0.8262$

図 1: The sound sources are human voice. (“い” and “え”).

5 おわりに

本稿では、集中マイクロホンの指向特性の相違を利用したブラインド音源分離を提案した。本手法は、マイクロホンの指向特性が未知であるようなブラインド状態での音源分離が可能である。

謝辞 本研究の一部は、日本学術振興会 21 世紀 COE プログラム「超高齢化社会における人とロボット技術の共生」の支援を受けて行われた。

参考文献

- [1] Matsumoto *et al.*, Multiple signal classification by aggregated microphones, IEICE Trans. Fundamentals, pp.1701-1707, Vol.E88-A, No.7, 2005.
- [2] Okada *et al.*, Direction measurement of a sound source in 3-D space utilizing unidirectional characteristics of microphone's sensitivity, Trans of SICE, pp.13-20, Vol.37, No.1, 2001.