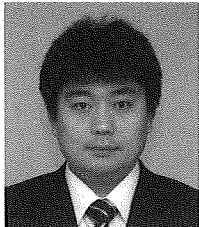


時々刻々と変動する音環境下での音源分離

Blind Source Separation under a Dynamic Acoustic Environment

1091002



研究代表者

熊本高等専門学校
情報通信エレクトロニクス工学科 准教授
(熊本電波工業高等専門学校 情報通信工学科 助 教)

石 橋 孝 昭

[研究の目的]

人間と機械のコミュニケーションにおいて、ユーザーに負担が無く、高速に情報伝達できるインターフェースは音声であり、多くの音声認識システムが実用化されている。しかし、現存の音声認識は、雑音のある環境では認識率が低下するため、雑音を含めて学習させる認識方法や、認識の前処理である雑音除去の研究が進められているが、音環境の変動で性能が低下する問題がある。

その原因の一つは、既存の音源分離では音源の情報は未知と設定しているにもかかわらず、音源数や音源の特徴が分かっている条件下で研究が進められているからである。このことが、実用化するときの大きな問題として立ちどころ、音源分離はなかなか実用化できない。そのため、音源分離の前処理として、これまでに固定された環境下で音源の数を推定する研究はいくつか報告されているが、処理時間がかかる問題がある。また、時々刻々と変動する音源数をリアルタイムで推定する研究の報告は見当たらない。さらに、音環境の変動を考慮した分離アルゴリズムに対する研究報告も無いようである。

そこで本研究では、音環境の変動と音源数を推定し、その結果に基づいて音源分離や雑音除去を行って、目的信号を抽出することのできるシステムの構築を目的とする。

[研究の内容, 成果]

1. ブラインド信号分離

N 個の音源 $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$ を M 個のマイクロホンで観測するとき、混合信号 $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$ は、混合行列 A によって

$$\mathbf{x}(t) = A\mathbf{s}(t) \quad (1)$$

のように表現できる。このとき、 $\mathbf{x}(t)$ と M のみが既知で、 $\mathbf{s}(t)$, N , A は未知である。

$\mathbf{x}(t)$ のみを用いて、 $\mathbf{s}(t)$ と A を推定することがブラインド信号分離の目的である。 A の逆関数 W を推定すれば、分離信号 $\mathbf{y}(t) = [y_1(t), \dots, y_N(t)]^T$ は

$$\mathbf{y}(t) = W\mathbf{x}(t) \quad (2)$$

のように得られる。

ブラインド信号分離問題を解く方法の一つに独立成分分析がある。独立成分分析は、混合信号が互いに統計的に独立な音源の重ね合わせであるという仮定の下で、混合信号から独立成分を分解する統計的手法であり、様々なアルゴリズムが提案されている。独立成分分析は音源のパワーに依存せず、非定常信号も分離できる特徴を持つ。

しかしながら、一般に、音源数 N はマイクロホン数 M と等しい仮定の下で導出されたア

ルゴリズムが多く、 $N \neq M$ のときに分離性能が低下するという問題がある。したがって、ブラインド信号分離を行うには、音源数を知るための前処理が必要となる。以降では、観測信号から音源数 N を推定し、目的音声を抽出する方法について報告する。

2. 変動する音環境下での音源分離

2.1 音環境の変動検出

音源は Fig. 1 のような話者音声で、観測信号は複数話者の同時発話となる場合を考える。また、二つのマイクロホン ($M=2$) による観測信号を用いる。

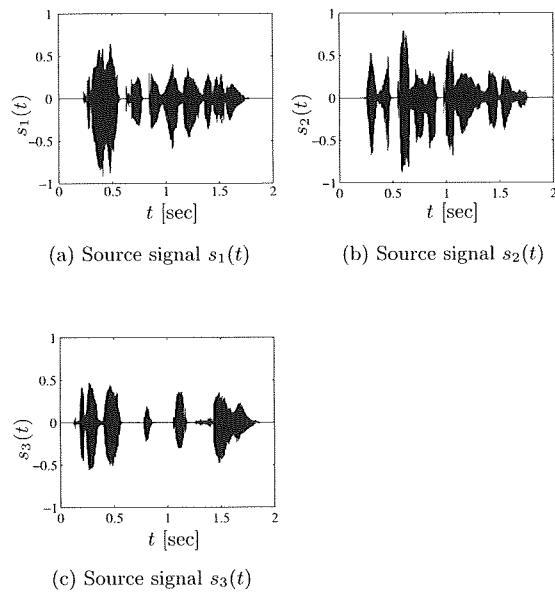


Fig. 1 source signals

まず、音源が存在しない場合、観測信号 $x_1(t)$ と $x_2(t)$ は共に振幅が 0 の信号となることは明らかである。したがって、全ての観測信号の振幅が 0 と判断できるときには、音源数は $N=0$ と容易に推定できる。

次に、音源が一つ (Fig. 1 の $s_1(t)$ のみが存在) の場合、観測信号 $x_1(t)$ と $x_2(t)$ は Fig. 2 のように得られる。観測信号は音源からマイクロホンまでの減衰率が異なるだけなので、振幅のスケールが異なる波形となる。

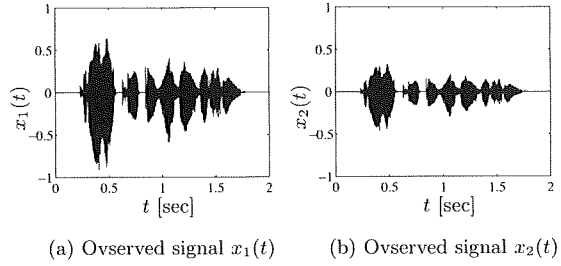


Fig. 2 Observed signals ($N=1$)

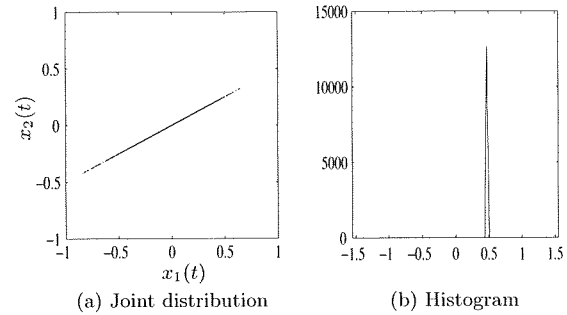


Fig. 3 Joint distribution and histogram ($N=1$)

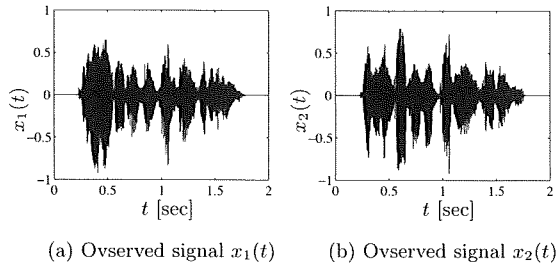
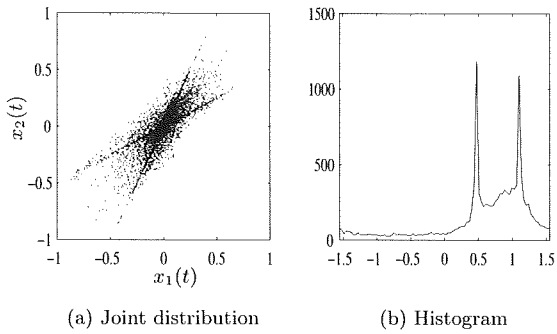
このときの観測信号の同時分布を Fig. 3 (a) に示す。グラフの横軸は $x_1(t)$ 、縦軸は $x_2(t)$ である。図から、スケールのみが異なるため同時分布は直線になることが分かる。また、この分布の方位に対するヒストグラムを Fig. 3 (b) に示す。すなわち、各観測時刻における角度 θ を

$$\theta(t) = \tan^{-1} \frac{x_2(t)}{x_1(t)} \quad (3)$$

を計算して、その角度に対するヒストグラムを作成する。横軸は $-\pi/2 \sim \pi/2$ [rad] の同時分布の方位、縦軸はその方位に対する頻度を示す。同時分布が一方向のみに存在するため、ヒストグラムは一つの尖ったピークを持つ。

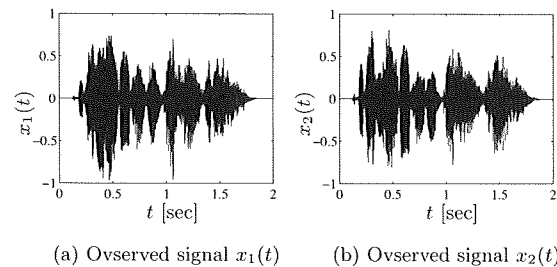
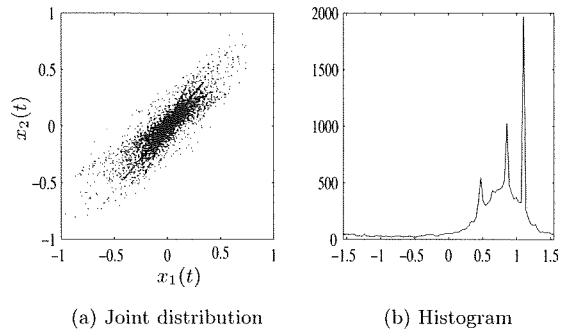
音源が二つ (Fig. 1 の $s_1(t)$ と $s_2(t)$ が存在) の場合、観測信号は Fig. 4 となる。複数の音源が混合したため、二つの観測信号の波形は全く異なる形状になることが分かる。

観測信号の同時分布を Fig. 5 (a)、そのヒストグラムを (b) に示す。(a) の同時分布から二本の直線上に多く分布していることが分


 Fig. 4 Observed signals ($N=2$)

 Fig. 5 Joint distribution and histogram ($N=2$)

かる。このことは、(b) のヒストグラムが二つの尖ったピークを持つことから明確である。

最後に、 $N=3$ (Fig. 1 の $s_1(t)$, $s_2(t)$, $s_3(t)$ が存在) の観測信号は Fig. 6 のように得られる。この同時分布を Fig. 7 (a) に、そのヒストグラムを (b) に示す。(a) の同時分布では特徴を見出すことが困難であるが、(b) のヒストグラムでは三つの尖ったピークの存在を確認できる。


 Fig. 6 Observed signals ($N=3$)

 Fig. 7 Joint distribution and histogram ($N=3$)

2.2 音源数推定

上述の事実から、観測信号の同時分布から作成されるヒストグラムは、音源数と等しい数の尖ったピークを持つことが分かる。したがって、このピーク数を求めることで音源数を推定できる。ピーク数を求めるために、本稿では、ヒストグラムの振幅 $h(k)$ の変動に着目し

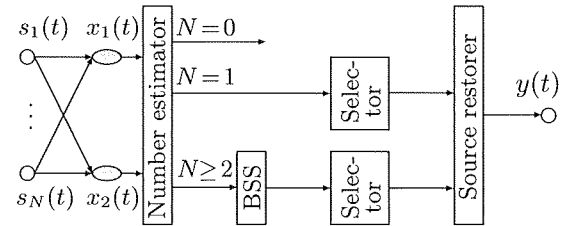


Fig. 8 Blind source separation based on the estimation for the number of the source signals

$$h(\kappa+1) - h(\kappa) \geq \varepsilon \quad (4)$$

を満たす点の数を求めることで、音源数を推定する。ここには、 κ はヒストグラムのフレーム番号、 ε は閾値である。

2.3 変動する音環境下での音源分離

音源数を推定した後、その数によって Fig. 8 のように音源分離や雑音除去の処理を行う。すなわち、音源数が 0 と推定された場合、何も出力しない。音源数が 1 と推定された場合、その音が目的音源と判断したときのみ出力する。音源数が 2 以上の場合、ブラインド信号分離によって音源分離し、分離記号から目的音源を選択し、出力する。以上の処理によって、変動する音響環境下での音源分離が実現可能となる。

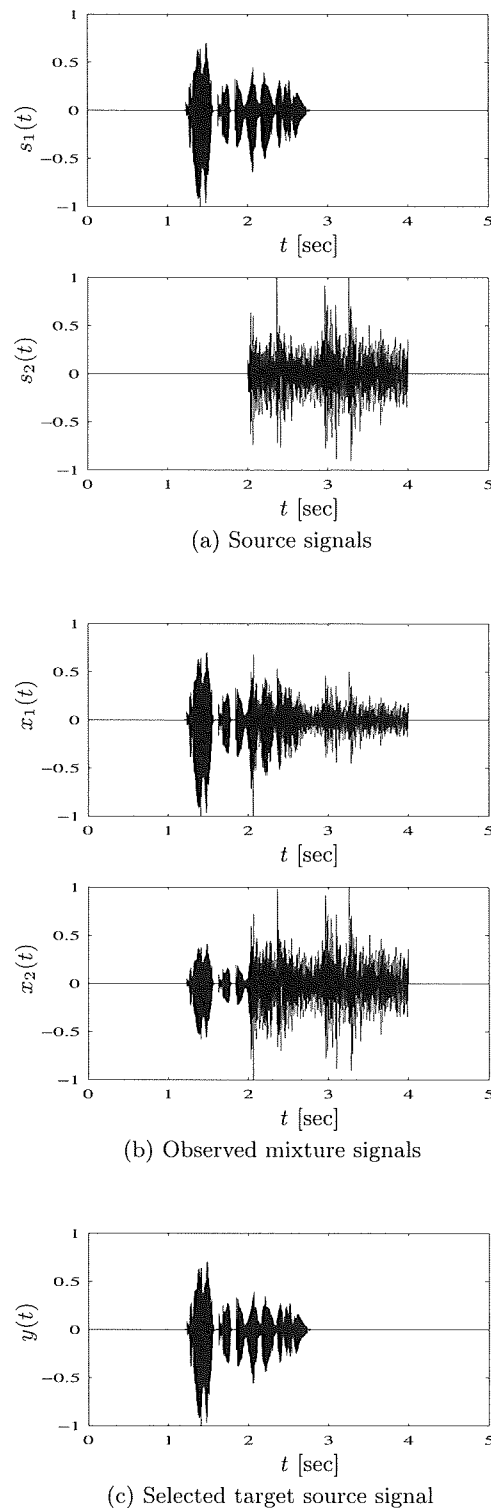


Fig. 9 Experimental results on blind source separation under a dynamic acoustic environment

3. 実験結果

本研究手法の有効性を検証するために実験を行った。音源は Fig. 9 (a) のように音源数が変動する音声を用いた。すなわち、混合信号の

約 0 ～ 1 秒は $N=0$ ，約 1 ～ 2 秒は $N=1$ で目的音声のみが存在し，約 2 ～ 3 秒は $N=2$ ，約 3 ～ 4 秒は $N=1$ で雑音のみが存在し，約 4 ～ 5 秒は $N=0$ とした。また，目的音声は $s_1(t)$ とした。これらの観測信号は (b) であり，観測信号から音源の波形や発話区間を判断することは困難である。観測信号を 0.5 秒ごとに分離した結果が (c) である。提案法は目的信号 $s_1(t)$ を抽出可能であり，その有効性が確認された。

[今後の研究の方向，課題]

音源分離や雑音除去に限らず，音響信号処理の分野における大きな問題に，高残響下で高精度に機能する処理方法の確立が挙げられる。本研究においても，高残響下で機能するアルゴリズムを開発することが今後の研究の方向である。また，本研究の実用化へ向けたアプリケーションの開発についても研究を進める予定である。

[成果の発表，論文等]

1. 石橋孝昭，井上勝裕，五反田博：ブラインド信号分離の基本的手法とその応用，システム制御情報学会誌，Vol. 54, No. 8 (2010)
2. T. Ishibashi, K. Inoue, H. Gotanda and K. Kumamaru: Frequency Domain Independent Component Analysis Without Permutation and Scale Indeterminacy, Proceedings of the 41 st ISCIE International Symposium on Stochastic Systems Theory and Its Applications, pp. 190 – 195 (2009)
3. 石橋孝昭，中島栄俊，岩崎宣生，五反田博：独立成分分析の初期値設定に関する考察，第 53 回システム制御情報学会研究発表講演会講演論文集，pp. 721 – 722 (2009)
4. 石橋孝昭，五反田博：音響信号のリアルタイムブラインド信号分離に対する試み，第 54 回システム制御情報学会研究発表講演会講演論文集，pp. 519 – 520 (2010)