

ブラインド条件での分離音の評価とパラメータ推定

6S-2

植田 護 橋本 周司
早稲田大学理工学部応用物理学科

1. はじめに

我々は、2つの音源によるモノラル競合音を各音源に固有な特徴情報を用いないブラインド状態で分離する問題について研究している。この問題をブラインドデコンポジション問題として捉えた一手法を、シミュレーションデータによる実験結果および実音の分離に応用した結果とともに以前に報告している[1] [2]。

しかし実音の分離に対して、この手法はある程度応用性があると思われるものの、システムのロバスト性に問題が残されていた。実音の場合には、雑音の影響や、音源に対する前提条件が完全に満たされない場合があり、推定されるシステムパラメータの信頼性に問題があると考えられる。そこで本稿では、ブラインド条件下で分離結果の信頼度を評価し、システムパラメータ候補の中から最適なものを選定するアルゴリズムを提案する。

2. 分離アルゴリズムの概要と問題点

前回提案した音源分離のためのブラインドデコンポジションアルゴリズムでは、以下のような条件を前提としていた。

前提条件(1) 混合音のパワースペクトルは、各音源のパワースペクトルの和で表される。

前提条件(2) 各音源の短時間パワースペクトルは、時間的に異なる変化をするが、スペクトルの形は、変化しない。

前提条件(3) 2つの音源のパワースペクトルの分布帯域が重ならない周波数 ω_0 が存在する。これらを式で表せば

$$F(t_n, \omega) = U(\omega) + V(\omega) \quad (1)$$

$$F(t_{n+1}, \omega) = a(t_{n+1}) \cdot U(\omega) + b(t_{n+1}) \cdot V(\omega) \quad (2)$$

($a(t_{n+1}) \neq b(t_{n+1})$)

ただし、 $\{U(\omega_0) \neq 0, \cap V(\omega_0) = 0\}$

または、 $\{V(\omega_0) \neq 0, \cap U(\omega_0) = 0\}$

$U(\omega), V(\omega)$: 2つの音源のパワースペクトル

$a(t_n), b(t_n)$: 各音源のパワースペクトルの時間変動

$F(t_n, \omega)$: 混合音パワースペクトル

ここで ω_0 の周波数は既知である必要はない。

まず、この前提条件と混合音パワースペクトルの時間変動を用いて、システムパラメータである音源の時間変動 $a(t_n), b(t_n)$ を以下の式で求める。

$$a(t_n) = \max[F(t_{n+1}, \omega)/F(t_n, \omega)] \quad (3)$$

$$b(t_n) = \min[F(t_{n+1}, \omega)/F(t_n, \omega)] \quad (4)$$

一般に ω_0 の条件を満たす周波数は複数存在するが、前提条件が満たされている場合には、式(3)、(4)のように ω に独立した定数値をシステムパラメータとする事ができる。

しかし実音のように、各音源のパワースペクトルの時間変動が周波数に依存した揺らぎをもち、また、雑音の混入がある場合には前提条件が満たされないため、式(3)、(4)による値の信頼度は低い。

また、このアルゴリズムでは、パワースペクトルの時間変動をとる時間窓間隔を固定しているが、音源のパワーの時間変化の度合いによってその間隔を能動的に変化させる方が实际的である。例えば、音源のパワーの時間変化が緩やかな場合、その間隔が短ければ、時間変動が微少なものになり計算誤差を招く。また逆に、時間窓の間隔が長すぎる場合には、前提条件の(2)の成立が難しくなる。

3. ブラインド状態での分離結果の評価法

もし、分離結果の信頼度を評価することができれば、システムパラメータの補正の基準にできる。

各音源のパワースペクトルが既知の場合には、次のような自乗誤差によって分離結果を評価することができる。

$$SQE = \sum_{\omega} \{(U(\omega) - \overline{U(\omega)})^2 + (V(\omega) - \overline{V(\omega)})^2\} \quad (5)$$

$\overline{U(\omega)}, \overline{V(\omega)}$: 推定スペクトル

しかし、各音源のパワースペクトルが未知であるブラインド状態ではこの式は使えない。そこで、混合音スペクトル及びあるシステムパラメータによって分離された推定スペクトルのみから信頼度を評価する次の方法を検討する。

式(5)は、式(6)、(7)のように書くことができる。

$$(U(\omega) - \overline{U(\omega)})^2 + (V(\omega) - \overline{V(\omega)})^2 = E(t_n, \omega) + 2\{\overline{U(\omega)}V(\omega) + U(\omega)\overline{V(\omega)} - U(\omega)V(\omega)\} \quad (6)$$

$$E(t_n, \omega) = F(t_n, \omega)^2 - 2(\overline{U(\omega)} + \overline{V(\omega)})F(t_n, \omega) + \overline{U(\omega)}^2 + \overline{V(\omega)}^2 \quad (7)$$

ここで、 ω_a は、 $F(t_{n+1}, \omega)/F(t_n, \omega)$ の値がその最大値の近傍の値をとる ω の集合、 ω_b は、最小値の近傍の値をとる ω の集合とすると、周波数領域 $\omega = \omega_a$ では $V(t_n, \omega) \approx 0$ 、 $\omega = \omega_b$ では $U(t_n, \omega) \approx 0$ 、と考えられる。この領域では、式(6)の右辺の第2項は非負であり、分離が良好な場合には0に近くなる。従って、式(7)はこの周波数領域で推定スペクトルの自乗誤差とみなせる。ここでは、次のような評価値によってシステムパラメータ $a(t_n), b(t_n)$ に対する信頼度を評価することとした。

$$\text{評価値} : NE(t_n) = \left\{ \sum_{\omega \in \omega_a, \omega_b} E(t_n, \omega) \right\} / \text{num}(\omega_a + \omega_b) \quad (8)$$

$\text{num}(\omega_a + \omega_b)$: ω_a, ω_b の総数

4. システムパラメータの自動補正

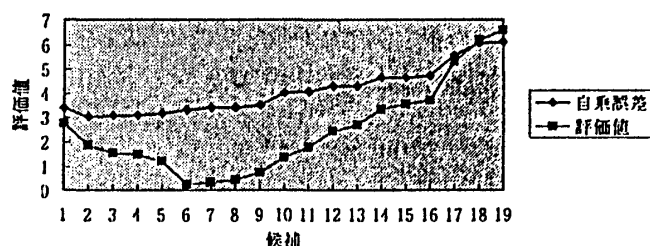
従来のアルゴリズムでは、混合音のパワースペクトルの変動 $F(t_{n+1}, \omega)/F(t_n, \omega)$ の最大値を $a(t_n)$ 、最小値を $b(t_n)$ として推定していたため、それらの値を実際の値よりも前者では過大に、後者では過小に評価することが多かったと考えられる。そこで、 $F(t_{n+1}, \omega)/F(t_n, \omega)$ の値の大きい順に $a(t_n)$ の候補、小さい順に $b(t_n)$ の候補として、それぞれの場合での分離結果の評価値を求める。それらのうち最も信頼度の高いシステムパラメータの組を選定する事とした。また、2つの時間窓の間隔をある時間単位で順に長くし、各時間間隔において最適なシステムパラメータを選定する。そのなかで、最も信頼度が高かった時間間隔での分離結果を最終的な結果とする。

5. 実験

今回、システムパラメータの候補は、混合音パワースペクトルの変動差の20%。時間窓は、0.1秒間隔で1.0秒まで変化させた。

5.1 時間窓間隔固定での評価値と自乗誤差の比較

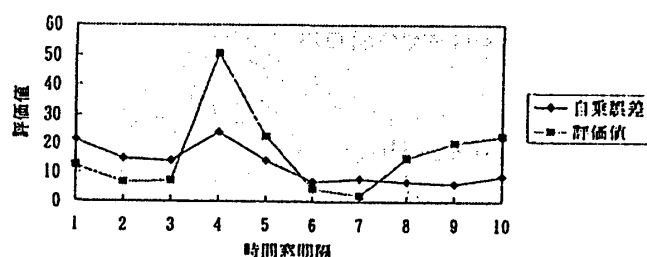
図[1]は、ある実音に対し、時間窓間隔を約0.1秒に固定し、音源の時間変動の各候補での式(8)と(5)の自乗誤差を比較したものである。自乗誤差の計算には、個別に計測した各音源のパワースペクトルを用いている。本稿での評価値が推定するパラメータ候補には、多少のズレがあるが、従来の方式と比べて改善していることがわかる。



図[1]

5.2 時間窓間隔を考慮に入れたパラメータ推定

図[2]は、ある実音に対しそれぞれの時間窓間隔における最良な評価値を算出し、そのときの推定スペクトルの自乗誤差とともに示している。ここでも、評価値と自乗誤差の間に強い相関が認められる。



図[2]

あとがき

本稿での評価値は、すべての周波数成分に対する推定スペクトルの評価を行うものではないが、この評価値を基に分離結果を最適化する事に成功し、従来の分離法に比べ、結果に誤差が生じる割合が減少する事が確かめられた。

【参考文献】・[1]植田, 橋本, 大照:「ブラインドデコンポジションとしての音源の分離と同定」: 情報処理学会, 第48回(H'5 前期) 全国大会予講集 1-pp389, 390(4T-5)
・[2]M.Ueda, S.Hashimoto, "Blind Decomposition of Concurrent Sounds": ICMC Proceedings 1994 pp.311-318