

4T-5

# ブラインドデコンポジション問題としての 音源の分離と同定

植田 護 橋本 周司 大照 完

早稲田大学理工学部応用物理学科

## 1. はじめに

音源分離問題は、複数音源の音響信号が重なり合った信号から、各音源、または、注目する特定音源の音響信号を抽出する問題である。しかし、その求解は、一般に非線形の逆問題となり、一意な解を求める事は困難である。このような問題に対しては、適当な制約条件を設定し、その有効な解法を見いだす必要がある。

人間は、混合音を片耳で受聴した場合でも分離認識できる。また、音源が音声や楽器音のように倍音構造を持つものに限らず、周波数成分が連続スペクトルで音階を持たない音源の場合や、初めて聞く音源、つまり事前情報のない音源が含まれている場合でも分離認識できる。これらのことは、人間が音源についての事前情報を持たないブラインド状態で、音源分離が可能であることを示唆している。我々は、2つの音源によるモノラルの競合音のパワースペクトルを、事前情報のないブラインド状態で分離する問題について検討している。ここでは、新たな制約条件を導入し、上に述べた人間の音源分離の特徴により近い性質を有する現実的な手法を実験結果とともに報告する。

## 2. Blind Decomposition 問題について

$n$ 個の音源  $A_i$  ( $i=1, 2, \dots, n$ ) があって、それらの音波の実時間波形を  $F_i(t)$  とするときそれらの音波の混合音波形  $D(t)$  は、

$$D(t) = F_1(t) + F_2(t) + \dots + F_n(t)$$

で表される。ここで、音源分離問題は、混合音  $D(t)$  から各音源の波形  $F_i(t)$  を求める逆問題となる。ここで  $F_i(t)$  について事前情報がある場合には、パターンマッチングなどの手法によって混合音  $D(t)$  を分離 (decompose) する事ができよう。

しかし、 $F_i(t)$  についての事前情報がない場合には、混合音  $D(t)$  の情報のみから  $F_i(t)$  を

分離する、「ブラインドデコンポジション問題」となる。一般に、完全なブラインド状態では、一意な解を求めることは不可能であるが、ブラインド状態を各音源の固有のモデルがないという事に限り、それぞれの音源に共通な性質を制約条件として工夫するならば、求解が可能である。人間の聴力が、極めてブラインドに近い状態で、音源分離が可能なる事から、ブラインドデコンポジション問題として音源分離問題を考察する事は、ヒューマンインターフェースの問題としても興味深い。

## 3. 音源に対する制約条件

ブラインド状態で、2音源の音源分離問題を解くための制約条件として以下の諸条件を設定する。

**制約条件(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。

**制約条件(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。

**制約条件(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  $\omega_0$  が存在する。これらを式で表せば、

$$Sh(t, \omega) = a(t) \cdot Sf(\omega) + b(t) \cdot Sg(\omega) \quad \text{式}[1]$$

$$(a(t) \neq b(t))$$

ただし、 $|Sf(\omega_0)| \neq 0$ , かつ  $Sg(\omega_0) = 0$

または、 $|Sg(\omega_0)| \neq 0$ , かつ  $Sf(\omega_0) = 0$

ここで、 $Sf(\omega), Sg(\omega)$  は、2つの音源のパワースペクトル、 $a(t), b(t)$  は、各音源のパワースペクトルの時間変動を表す。 $Sh(t, \omega)$  は、混合音パワースペクトルを表す。

制約条件(1)は、音源が純音の場合には成立が困難であるが、複雑な音源波形の場合には、両音源の相互相関は小さくなり近似的に成立すると思われる。制約条件(2)は、変数分離条件で、 $Sf(\omega), Sg(\omega), a(t), b(t)$  は、未知である。また、パワースペクトルの時間変動が周波数によらず一定である事は、一般の場合において成り立つとは言えないが、スペクトルをとる間隔が短ければ、ある程度仮定できる。制約条件(3)の必要性は後に述べるが、この条件を満たす周波数  $\omega_0$  の値は既知である必要はない。

### 4. アルゴリズム

式[1]において、 $Sh(t, \omega)$  から各音源パワースペクトル  $Sf(\omega), Sg(\omega)$  を求めることが目的である。離散化信号の時間区間 ( $t_n \sim t_{n+\Delta t}$ ) の時間窓長でのパワースペクトルを時刻  $t_n$  でのパワースペクトルと呼ぶ。ここで、時刻  $t_0$  から  $t_1$  の各音源の時間変動を  $a, b$ , と表せば、式[1]は、 $a \neq b$  として

$$Sh(t_0, \omega) = Sf(\omega) + Sg(\omega) \quad \text{式[2]}$$

$$Sh(t_1, \omega) = a \cdot Sf(\omega) + b \cdot Sg(\omega) \quad \text{式[3]}$$

として表される。ここで、正の実数  $\rho$  をパラメータとして、観測される  $Sh(t_0, \omega), Sh(t_1, \omega)$  できまる関数  $K(\rho, \omega)$  を定義する。

$$K(\rho, \omega) = Sh(t_1, \omega) - \rho \cdot Sh(t_0, \omega) \quad \text{式[4]}$$

ここで、式[2], [3] を用いると、

$$K(\rho, \omega) = (a - \rho) \cdot Sf(\omega) + (b - \rho) \cdot Sg(\omega) \quad \text{式[5]}$$

と表される。 $Sf(\omega), Sg(\omega)$  は、パワースペクトルであるから非負である。ここで、 $a < b$  として考え、 $\rho$  を0から動かす。 $\rho < a$  の間は、 $K(\rho, \omega)$  は、すべての周波数において正であるが、 $\rho$  が  $a < \rho < b$  の領域に入るとき制約条件(3)を満たす周波数  $\omega_0$  において、 $K(\rho, \omega_0) < 0$  となる  $\omega_0$  が出現する。すなわち式[4]において  $\rho$  の値を0から大きくしていき、負となる  $K(\rho, \omega)$  を与え始める  $\rho$  が、求める  $a$  の値である。また、 $b < \rho$  の場合、 $K(\rho, \omega)$  は全ての周波数成分において負になることから、同様にして、 $b$  の値も求めることができる。

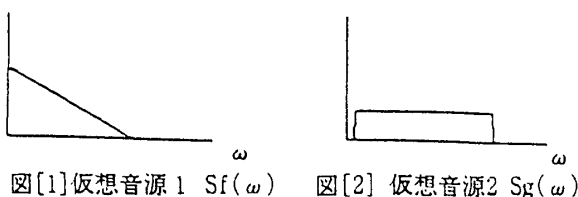
このようにして、 $a, b$  の値が  $K(\rho, \omega)$  から、求まり、これを式[3]に代入すれば、式[2] とから連立方程式を解いて、 $Sf(\omega), Sg(\omega)$  を求める。

### 5. 実験

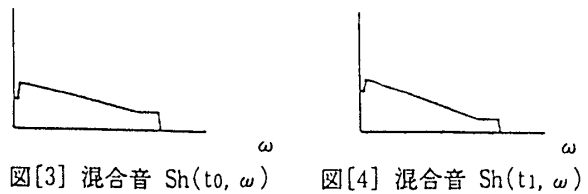
以上に述べたアルゴリズムを用いた音源分離システムを構築し、本稿で設定した制約条件を満たすスペクトルモデルを用いたシミュレーション分離実験と、実音による分離実験を行った。システムは、入力をモノラルで、12bit, 20KHzサンプリングにより、AD変換し、2048点をブロックとしてFFT解析を行った。

#### 5.1 シミュレーションデータを用いた実験

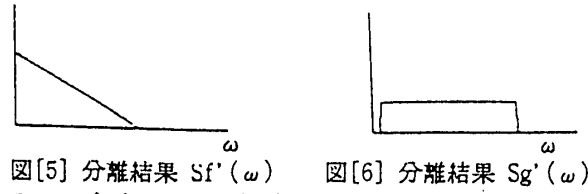
図[1], [2] に示すような連続スペクトルモデルに計算機上で時間変動を与えて合成したものが、図[3], [4] である。この2つを分離システムに入力し、その出力が、図[5], [6] である。



図[1] 仮想音源1  $Sf(\omega)$     図[2] 仮想音源2  $Sg(\omega)$



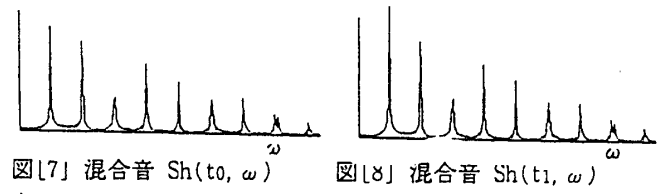
図[3] 混合音  $Sh(t_0, \omega)$     図[4] 混合音  $Sh(t_1, \omega)$



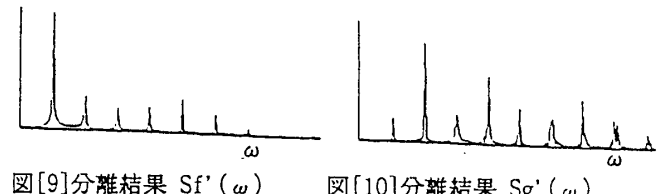
図[5] 分離結果  $Sf'(\omega)$     図[6] 分離結果  $Sg'(\omega)$

### 5.2 実音による実験

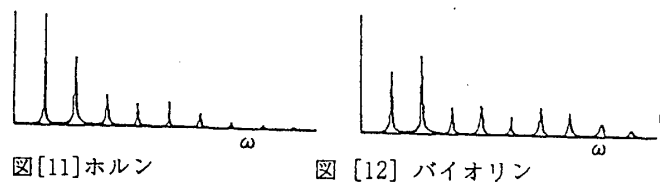
図[7], [8] は、ホルンとバイオリンを同一音高で、同時に一つのマイクで集音したもののパワースペクトルである。これを分離システムに入力した結果が、図[9], [10]である。ホルン、バイオリン、それぞれの単独でのパワースペクトル図[11], [12]と比較すると、それぞれの特徴が分離されていることが判る。



図[7] 混合音  $Sh(t_0, \omega)$     図[8] 混合音  $Sh(t_1, \omega)$



図[9] 分離結果  $Sf'(\omega)$     図[10] 分離結果  $Sg'(\omega)$



図[11] ホルン    図[12] バイオリン

### 6. あとがき

このほかに、2つの母音「あ」と「い」を同時に発音したものに対しても分離実験を行い、その分離スペクトルを実時間信号に逆変換した音を実際に聞き比べてみたところ、ほぼ納得のいく結果が得られた。今後は、本方式のロバスト性、3つ以上の音源に対する応用などの問題点について検討しつつ、会話システムなどへの応用を考えていきたい。

#### 【参考文献】

- ・植田, 橋本, 大照;"スペクトル変動を用いた音源分離の試み", 情報処理学会, 第46回(H'5 前期) 全国大会
- ・柏野;"聴覚における音響事象の分離と補完" 日本音響学会聴覚, H-92-20, 1992