

調波成分分析によるブラインド音楽信号分離の検討

小野 順貴[†] 細谷 弘[†] 和泉 洋介[†] 嵯峨山茂樹[†]

[†] 東京大学大学院 情報理工学系研究科 システム情報学専攻

〒 113-8656 東京都文京区本郷 7-3-1

E-mail: †{onono,h-hosoya,izumi,sagayama}@hil.t.u-tokyo.ac.jp

あらまし 本研究では、音楽信号を対象としたブラインド音源分離を目的とした調波成分分析という枠組みについて論じる。ブラインド音源分離に広く用いられている独立成分分析においては、音源信号間の独立性を仮定し音源分離を実現するが、基本的には音源信号数以上の観測信号が必要である。一方、対象が音楽信号の場合には、音源信号は定常な音高の連なりであることが多く、各時間フレームで調波性であることを仮定できると考えられる。調波成分分析とはこの性質に着目し、音源信号の各時間フレームにおける調波性を最大化することによってブラインド音源分離を行うことを目指すものであり、調波性という、独立性より強い仮定を導入することによって、劣決定の条件下でも有効に働くことが期待できる。本稿では、観測信号の二次形式で表される調波性の評価指標に基づく定式化と、基礎的なシミュレーション実験の結果について報告する。

キーワード 調波成分分析, 独立成分分析, ブラインド音源分離, 音楽信号, 調波性

Blind Separation of Music Signals by Harmonic Component Analysis

Nobutaka ONO[†], Hiroshi HOSOYA[†], Yosuke IZUMI[†], and Shigeki SAGAYAMA[†]

[†] Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo

7-3-1 Hongo Bunkyo-ku Tokyo 113-8656, Japan

E-mail: †{onono,h-hosoya,izumi,sagayama}@hil.t.u-tokyo.ac.jp

Abstract In this paper, we discuss Harmonic Component Analysis (HCA) as a framework for blind source separation of music signals. In Independent Component Analysis (ICA), which has been widely used for blind source separation, the separation is realized based on the assumption that source signals are mutually independent, where more observation signals than source signals are required. While, in the case of music signals, not independency but harmonicity on each time frame can be assumed because they consists of tones with stable pitches. Focusing on the properties, HCA aims to separate sources by maximizing harmonicity of sources, which will be effective for underdetermined mixtures due to introducing harmonicity, which is stronger assumption than independency. In this paper, we present formulation based on the harmonicity criterion represented quadratic form of observation signals, and results of fundamental simulative experiments.

Key words Harmonic Component Analysis, Independent Component Analysis, blind source separation, music signals, harmonicity

1. はじめに

音楽信号を対象とした音源分離は、音楽検索、加工、自動採譜といった様々な応用の基礎技術として有用であり、特に近年、活発に研究が進められている [1] ~ [4]。CD に記録するためにステレオ信号に remix された録音や、複数マイクロフォンで録音されたライブ録音など、多くの場合においてその混合過程

は未知であり、観測信号のみからこうした分離を行うブラインド分離が求められる。ブラインド信号分離に対しては独立成分分析 [5] が標準的な手法として確立しているが、音楽信号の場合には、音源信号数が観測信号数よりも多い劣決定な状況がしばしば発生するため、そのままでは直接適用できない場合も多く存在する。そもそも劣決定の条件下では、混合過程が既知であったとしても解を一意に定めることはできないため、音源信

号にさらに強い仮定を必要とする。例えば音声信号の場合には、時間周波数領域における音源信号のエネルギー分布がスパースであり、互いに重ならないことがよく仮定されるが[6]~[8]、音楽信号の場合には、音源信号のエネルギー分布が疎であったとしても、基本周波数は音階を構成する半音単位にほぼ離散化されており、さらに音源信号の高調波同士を互いに重ねあうことで和音を構成することが多いため、この仮定は必ずしも妥当ではない状況が生じる。

一方、音楽信号は多くの場合、短い時間フレームでは安定した基本周波数をもつ調波信号であるとみなせることが多い。本研究はこの点に着目し、音楽音響信号を対象に、音源信号の調波性を最大化することによりブラインド音源分離を行う、調波成分分析という新たな枠組みを論じる。調波性自体は従来から、計算論的聴覚情景解析や正弦波重畳モデルの枠組みなどに積極的に用いられているが[9]~[13]、単一信号の分離の場合には時間方向のパーミュテーション問題を解く必要が生じる。これに対し、劣決定ではあっても複数の観測信号に対して調波性を積極的に利用することにより、1) 強度比、時間差の情報による時間方向のパーミュテーション問題の解決、2) 調波性というスペクトル構造の仮定による周波数方向のパーミュテーション問題の解決が期待できる。

本稿では、調波性の定量的な指標の与え方と、これに基づく劣決定条件下での分離の原理を中心に、理論的な定式化と基礎的な評価実験の結果について報告する。

2. 調波成分分析の理論

2.1 問題設定

いま、短時間フーリエ変換により時間周波数領域で表現された観測信号と音源信号を $O_{j,\omega,t}$, $S_{i,\omega,t}$ により表す。ここで i, j はマイクロフォンと音源のインデックス、 ω, t は、周波数と時間フレームのインデックスである。音源 j からマイクロフォン i までのインパルス応答長に対し短時間フーリエ変換のフレーム長が十分長ければ、音源信号と観測信号の関係は近似的に、

$$O_{\omega,t} = A_{\omega} S_{\omega,t} \quad (1)$$

のように表される。ただし $S_{\omega,t}$ と $O_{\omega,t}$ は

$$S_{\omega,t} = (S_{1,\omega,t}, \dots, S_{K,\omega,t})^T \quad (2)$$

$$O_{\omega,t} = (O_{1,\omega,t}, \dots, O_{L,\omega,t})^T \quad (3)$$

により定義される音源信号、観測信号の時間周波数毎のベクトル表現であり K と L はそれぞれ、音源とマイクロフォンの個数を表す。ブラインド音源分離の問題とは、混合行列 A_{ω} が未知という条件の下、観測信号 $O_{i,\omega,t}$ のみから $S_{j,\omega,t}$ を推定することである。

いま、全時間フレーム数を M 、全周波数 bin 数を N とすると、観測によって得られる変数 $O_{j,\omega,t}$ は LMN 個、一方未知数は、混合行列の全成分数 KLN 個と音源信号を表す $S_{i,\omega,t}$ が KMN 個の和であるから、 $LMN < KLN + KMN$ 、すなわち、

$$\frac{LM}{L+M} < K \quad (4)$$

の条件では、式 (1) だけから解を一意に決めることはできない。また、式 (1) の右辺は 2 つの未知数の積であるから、定数倍の任意性が含まれる。(ある A_{ω} , $S_{i,\omega,t}$ が解であれば、0 でない任意の定数 C に対して CA_{ω} , $(1/C)S_{i,\omega,t}$ も解となる。) よってブラインド音源分離を解くためには、式 (1) を満たす組 $(A_{\omega}, S_{\omega,t})$ の中から解を決定するための別の評価基準が必要であり、その中には定数倍の任意性の解消が含まれている必要がある。例えば独立成分分析においては音源信号間の独立性、スパース性を利用した劣決定のブラインド音源分離においては、音源方向クラスタ内の分散や L_1 ノルムなどが評価関数となっている。以下では、調波性の評価指標と定数倍の任意性の解消について順に論じる。

2.2 調波信号の確率モデル

調波成分分析は、各時間フレームにおける音源信号の調波性ができるだけ大きくなるようにすることで音源分離を試みるものであり、これは調波信号の確率モデルに基づく最尤法として定式化できる。調波信号が従う確率密度分布のモデルとしては様々なものが考えられるが、本研究ではそもそもの調波信号(周期信号)の定義: $f(t) = f(t+T)$ (T : 基本周期)に着目し、この定義をフィッティングした際の残差パワー:

$$\int |f(t) - f(t+T)|^2 dt \quad (5)$$

に基づくモデル化を試みる。

まず、時間周波数表現された音源信号 $S_{i,\omega,t}$ の尤度を計算しやすいうように、式 (5) を Parseval の等式を用いて周波数領域で表すと、

$$\begin{aligned} \int |f(t) - f(t+T)|^2 dt &= \frac{1}{2\pi} \int |\mathcal{F}[f(t) - f(t+T)]|^2 d\omega \\ &= \frac{1}{2\pi} \int |(1 - e^{j\omega T})F(\omega)|^2 d\omega \\ &= \frac{2}{\pi} \int \sin^2 \frac{\omega T}{2} |F(\omega)|^2 d\omega \end{aligned} \quad (6)$$

を得る。よって時間フレーム t における音源 j に対する式 (5) の残差パワーは、

$$Q_{j,\omega,t} = \sin^2 \frac{\omega T_{j,t}}{2} \quad (7)$$

という周波数特性をもった楕円ノッチフィルタに通した出力パワー:

$$P_{j,t} = \sum_{\omega} |Q_{j,\omega,t} S_{j,\omega,t}|^2 \quad (8)$$

とみなせる。ただし $T_{j,t}$ は、時間フレーム t での音源 j の基本周期であり、 ω に対する和は、負の周波数を含めた全ての周波数 bin に対してとるものとする。理想的な調波信号であれば $P_{j,\omega,t}$ は 0 となるから、 $P_{j,\omega,t}$ が、各音源、各時間フレーム毎に独立に平均 0、分散 σ_j^2 のガウス分布に従うと仮定すれば、全音源・時間周波数での $S_{j,\omega,t}$ の対数尤度は

$$\begin{aligned}
\log p(\mathbf{S}) &= \log \prod_{j,\omega,t} p(S_{j,\omega,t}) \\
&= \sum_{j,\omega,t} \log p(S_{j,\omega,t}) \\
&= \sum_{j,\omega,t} \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \sigma_j^2 - \frac{|Q_{j,\omega,t} S_{j,\omega,t}|^2}{2\sigma_j^2} \right\}
\end{aligned} \tag{9}$$

と表わされる。ここで (i, j) 成分が

$$(Q_{\omega,t})_{(i,j)} = \begin{cases} \frac{1}{\sigma_j} \sin \frac{\omega T_{j,t}}{2} & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{10}$$

と定義される対角行列 $Q_{\omega,t}$ を用い、定数項を省略し、全体を2倍すると、対数尤度は

$$J = -MN \sum_j \log \sigma_j^2 - \sum_{\omega,t} \mathbf{S}_{\omega,t}^h Q_{\omega,t}^2 \mathbf{S}_{\omega,t} \tag{11}$$

と表わされる。ただし、 $Q_{j,\omega,t}$ に含まれる基本周期 $T_{j,t}$ は未知であるから、 $T_{j,t}$ もこの対数尤度の最大化によって決める必要があることに注意する。

2.3 定数倍の曖昧性の解消

式 (11) は明らかに $S_{\omega,t}$ のノルムが小さいほど大きい値をとるので、 $S_{\omega,t}$ のスケールを拘束する必要がある。独立成分分析では、前処理として観測信号の分散を1に規格化した後に、分離行列 $W_\omega = A_\omega^{-1}$ をエルミート行列に限ることで、音源信号の分散を1に規格化することがよく行われるが、劣決定の場合には、この手法は適用できない。一方、スパース性を利用した劣決定のブラインド音源分離では、混合行列の列ベクトルの成分の1つを1に固定されることが多い。ただしこの場合には、ある特定の観測信号を他に対して特別扱いしたことになるので、対称性が悪い面もある。よってここでは、 A_ω の列ベクトルのノルムを1に規格化することでスケールを決めることとした。 e_i を i 番目の成分のみが1で他の成分が0である K 次元のベクトルとするとこの条件は、 $1 \leq \omega \leq N, 1 \leq i \leq K$ に対して、

$$e_i^h A_\omega^h A_\omega e_i = 1 \tag{12}$$

と表わされる。

3. 更新式の導出

3.1 ラグランジュ未定乗数法による目的関数の設定

前節での定式化によりここでの問題は、式 (1) の観測モデルと式 (12) の制約条件の下で、式 (11) を最大化するパラメータ $S_{\omega,t}, A_\omega, T_{j,t}$ を求めることである。ラグランジュ未定乗数を導入すれば、

$$\begin{aligned}
J &= -MN \sum_j \log \sigma_j^2 - \sum_{\omega,t} \mathbf{S}_{\omega,t}^h Q_{\omega,t}^2 \mathbf{S}_{\omega,t} \\
&\quad + \sum_{\omega,t} \lambda_{\omega,t} (\mathbf{O}_{\omega,t} - A_\omega \mathbf{S}_{\omega,t}) \\
&\quad + \sum_{\omega} \sum_i \lambda_{\omega,i} (e_i^h A_\omega^h A_\omega e_i - 1)
\end{aligned} \tag{13}$$

と表される。

直接解は求まらないため、各変数を固定して逐次的に更新する反復推定のための更新則を導出する。

3.2 音源信号 $S_{\omega,t}$ の更新

$S_{\omega,t}$ による式 (13) の偏微分は、

$$\frac{\partial J}{\partial S_{\omega,t}} = -Q_{\omega,t}^2 \mathbf{S}_{\omega,t} - A_\omega^h \lambda_{\omega,t} \tag{14}$$

と表わされるので、これを0とおくと、

$$\mathbf{S}_{\omega,t} = -Q_{\omega,t}^{-2} A_\omega^h \lambda_{\omega,t}$$

を得る。式 (1) に代入すると、

$$\mathbf{O}_{\omega,t} = -A_\omega (Q_{\omega,t}^{-2} A_\omega^h \lambda_{\omega,t}) \tag{15}$$

すなわち、

$$\lambda_{\omega,t} = -(A_\omega Q_{\omega,t}^{-2} A_\omega^h)^{-1} \mathbf{O}_{\omega,t} \tag{16}$$

よって、

$$\mathbf{S}_{\omega,t} = Q_{\omega,t}^{-2} A_\omega^h (A_\omega Q_{\omega,t}^{-2} A_\omega^h)^{-1} \mathbf{O}_{\omega,t} \tag{17}$$

を得る。

3.3 混合行列 A_ω の更新

式 (13) で表される目的関数は、陽には $S_{\omega,t}$ の関数であり、 A_ω は制約条件を介して含まれている。いま、 $S_{\omega,t}$ の最尤値として求めた式 (17) を式 (13) に代入すると、

$$\begin{aligned}
\tilde{J} &= - \sum_t \mathbf{O}_{\omega,t}^h (A_\omega Q_{\omega,t}^{-2} A_\omega^h)^{-1} \mathbf{O}_{\omega,t} \\
&\quad + \sum_{\omega} \sum_i \lambda_{\omega,i} (e_i^h A_\omega^h A_\omega e_i - 1) - MN \sum_j \log \sigma_j^2
\end{aligned} \tag{18}$$

のように、 A_ω を陽に含む形を得る。 \tilde{J} の A_ω による偏微分は

$$\begin{aligned}
\frac{\partial \tilde{J}}{\partial A_\omega} &= - \sum_t (A_\omega Q_{\omega,t}^{-2} A_\omega^h)^{-1} \mathbf{O}_{\omega,t} \mathbf{O}_{\omega,t}^h (A_\omega Q_{\omega,t}^{-2} A_\omega^h)^{-1} A_\omega Q_{\omega,t}^{-2} \\
&\quad + A_\omega D_\omega
\end{aligned} \tag{19}$$

となる。ただし、 D_ω は (i, j) 成分が

$$(D_\omega)_{(i,j)} = \begin{cases} \lambda_{\omega,j} & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{20}$$

で表される、ラグランジュ未定乗数を成分とする対角行列である。式 (19) には、 A_ω が複雑な形で含まれており、これを満たす A_ω を解析的に求める式は導出できていない。よってここでは、 A_ω の更新は勾配法を用いて行った。具体的には、

$$\begin{aligned}
\Delta A_\omega &= -\beta \sum_t (A_\omega Q_{\omega,t}^{-2} A_\omega^h)^{-1} \mathbf{O}_{\omega,t} \mathbf{O}_{\omega,t}^h (A_\omega Q_{\omega,t}^{-2} A_\omega^h)^{-1} A_\omega Q_{\omega,t}^{-2}
\end{aligned} \tag{21}$$

を用いて $A_\omega \leftarrow A_\omega + \Delta A_\omega$ と更新を行い、その後各列ベクトルのノルムを1に規格化し直すことで、式 (12) の制約条件の下で A_ω を更新した。ただし β は、更新のステップ幅を決める正の実数である。

3.4 基本周期 $T_{j,t}$ の更新

式 (11) には, $T_{j,t}$ は $Q_{j,\omega,t}$ の中に含まれている。いま式 (11) の中で, ある $T_{j,t}$ を含む項をまとめたものを $J_{T_{j,t}}$ とおき, $T_{j,t}$ はサンプリング周期の整数倍と仮定すると,

$$\begin{aligned} J_{T_{j,t}} &= - \sum_{\omega} \sin^2 \frac{\omega T_{j,t}}{2} |S_{j,\omega,t}|^2 \\ &= - \frac{1}{2} \sum_{\omega} (1 - \cos \omega T_{j,t}) |S_{j,\omega,t}|^2 \\ &= - \frac{1}{2} \sum_{\omega} |S_{j,\omega,t}|^2 + \frac{1}{2} \sum_{\omega} e^{j\omega T_{j,t}} |S_{j,\omega,t}|^2 \quad (22) \end{aligned}$$

を得る。ここで ω に対する和は負の周波数を含めたすべての周波数 bin に対してとっており, 偶関数である $|S_{j,\omega,t}|^2$ と奇関数である $\sin \omega T_{j,t}$ の積和が 0 になることを用いた。 $J_{T_{j,t}}$ を最大とする $T_{j,t}$ は解析的には求まらないが, 式 (22) の第 2 項は時間フレーム j における自己相関関数であり, これが最大となる $T_{j,t}$ を離散値全探索 (探索範囲は, 対象とする信号として起こりえる範囲) によって行うことができる。また実際には, 最大ピーク近傍 3 点で二次関数フィッティング等により, $T_{j,t}$ をサブサンプルまで求める必要がある。

4. 基本周期を用いた劣決定分離の原理

本稿で提案する手法は, 前節で議論した更新式を用いて

- 混合行列 A_{ω} の更新
- 音源分離 ($S_{j,\omega,t}$ の更新)
- ピッチ (基本周期) 推定 ($T_{j,t}$ の更新)

を繰り返すことにより, 最尤の意味での音源分離を行うものであるが, 特に音源分離ステップでの性質を把握するために, 真の混合行列 A_{ω} が得られたものと仮定して $S_{j,\omega,t}$ の更新式である式 (17) に着目する。式 (17) は, 以下のような性質をもつ。

性質 1 式 (17) は, $Q_{\omega,t}^{-2}$ のスケールに依らない。すなわち, $Q_{\omega,t}^{-2}$ と $CQ_{\omega,t}^{-2}$ (C は 0 でない任意の実数) は同じ解を与える。

証明:

$$\begin{aligned} &(CQ_{\omega,t}^{-2})A_{\omega}^h(A_{\omega}(CQ_{\omega,t}^{-2})A_{\omega}^h)^{-1}\mathbf{O}_{\omega,t} \\ &= CQ_{\omega,t}^{-2}A_{\omega}^h(1/C)(A_{\omega}Q_{\omega,t}^{-2}A_{\omega}^h)^{-1}\mathbf{O}_{\omega,t} \\ &= Q_{\omega,t}^{-2}A_{\omega}^h(A_{\omega}Q_{\omega,t}^{-2}A_{\omega}^h)^{-1}\mathbf{O}_{\omega,t} \quad (23) \end{aligned}$$

より明らか。

性質 2 $Q_{\omega,t}^{-2}$ の (i, i) 成分が 0 のとき, 式 (17) が与える解において, i 番目の成分は 0 となる。

証明:

$Q_{\omega,t}^{-2}$ は対角行列であり, 式 (17) の右辺に左から乗じられているので明らか。

性質 3 $Q_{\omega,t}^{-2} = I$ のとき, 式 (17) はノルム最小解を与える。

証明:

$Q^{-2} = I$ を式 (17) に代入すると,

$$\begin{aligned} \mathbf{S}_{\omega,t} &= IA_{\omega}^h(A_{\omega}IA_{\omega}^h)^{-1}\mathbf{O}_{\omega,t} \\ &= A_{\omega}^h(A_{\omega}A_{\omega}^h)^{-1}\mathbf{O}_{\omega,t} \quad (24) \end{aligned}$$

である。一方, $\mathbf{O}_{\omega,t} = A_{\omega}\mathbf{S}_{\omega,t}$ を満たす解の中で $|\mathbf{S}_{\omega,t}|^2$ を最小とするものを求めると,

$$J_2 = \mathbf{S}_{\omega,t}^h \mathbf{S}_{\omega,t} + \lambda^h (\mathbf{O}_{\omega,t} - A_{\omega} \mathbf{S}_{\omega,t}) \quad (25)$$

を $\mathbf{S}_{\omega,t}$ で微分して 0 とおくことにより,

$$\mathbf{S}_{\omega,t} = A_{\omega}^h \lambda \quad (26)$$

すなわち,

$$\mathbf{O}_{\omega,t} = A_{\omega} A_{\omega}^h \lambda \quad (27)$$

$A_{\omega} A_{\omega}^h$ が正則であると仮定すれば,

$$\lambda = (A_{\omega} A_{\omega}^h)^{-1} \mathbf{O}_{\omega,t} \quad (28)$$

であるので, 式 (26) に代入し,

$$\mathbf{S}_{\omega,t} = A_{\omega}^h (A_{\omega} A_{\omega}^h)^{-1} \mathbf{O}_{\omega,t} \quad (29)$$

を得る。すなわち式 (24) は, $\mathbf{O}_{\omega,t} = A_{\omega} \mathbf{S}_{\omega,t}$ を満たす解の中で $|\mathbf{S}_{\omega,t}|^2$ を最小とするものであることがわかる。

性質 4

K 個の成分のうち, 0 でない成分が高々 L 個であるような音源信号 $\mathbf{S}_{\omega,t}$ に対する観測信号 $\mathbf{O}_{\omega,t} = A_{\omega} \mathbf{S}_{\omega,t}$ が得られたものとする。 $\mathbf{S}_{\omega,t}$ の 0 成分のインデックスの集合を $\Omega = \{j | S_{j,\omega,t} = 0\}$ とするとき, $\forall j \in \Omega$ に対して $Q_{\omega,t}^{-2}$ の (j, j) 成分が 0 であるならば, 式 (17) は正しい解を与える。

証明:

一般性を失わず, $\mathbf{S}_{\omega,t}$ の成分の順番を入れ替え, 0 でない成分をと 0 成分をそれぞれまとめて,

$$\mathbf{S}_{\omega,t} = \begin{pmatrix} \hat{\mathbf{S}}_{\omega,t} \\ \mathbf{0} \end{pmatrix} \quad (30)$$

と表わせる。ただし $\hat{\mathbf{S}}_{\omega,t}$ は L 個の成分をもつベクトルとする。これにあわせて混合行列 A_{ω} も列ベクトルを並べ替えた上でブロック化し,

$$A_{\omega} = \begin{pmatrix} A_1 & A_2 \end{pmatrix} \quad (31)$$

のように表せる (A_1 は $L \times L$, A_2 は $L \times K$ の行列)。このとき, 観測信号 $\mathbf{O}_{\omega,t}$ は,

$$\begin{aligned} \mathbf{O}_{\omega,t} &= A_{\omega} \mathbf{S}_{\omega,t} \\ &= \begin{pmatrix} A_1 & A_2 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{S}}_{\omega,t} \\ \mathbf{0} \end{pmatrix} \\ &= A_1 \hat{\mathbf{S}}_{\omega,t} \quad (32) \end{aligned}$$

と表せる。一方, $\forall j \in \Omega$ に対して $Q_{\omega,t}^{-2}$ の (j, j) 成分が 0 であることを仮定すれば,

$$Q_{\omega,t}^{-2} = \begin{pmatrix} Q_1 & O \\ O & O \end{pmatrix} \quad (33)$$

と表せる (Q_1 は $L \times L$ の行列)。式 (32), 式 (33) の下では,

$$\begin{aligned} & (A_{\omega} Q_{\omega,t}^{-2} A_{\omega}^h)^{-1} \\ &= \left\{ \begin{pmatrix} A_1 & A_2 \end{pmatrix} \begin{pmatrix} Q_1 & O \\ O & O \end{pmatrix} \begin{pmatrix} A_1^h \\ A_2^h \end{pmatrix} \right\}^{-1} \\ &= (A_1 Q_1 A_1^h)^{-1} \\ &= A_1^{-h} Q_1^{-1} A_1^{-1} \end{aligned} \quad (34)$$

であるので (ただし A_1 の正則性を仮定した), 式 (17) の右辺は,

$$\begin{aligned} & Q_{\omega,t}^{-2} A_{\omega}^h (A_{\omega} Q_{\omega,t}^{-2} A_{\omega}^h)^{-1} O_{\omega,t} \\ &= \begin{pmatrix} Q_1 A_1^h \\ O \end{pmatrix} (A_1^{-h} Q_1^{-1} A_1^{-1}) A_1 \hat{S}_{\omega,t} \\ &= \begin{pmatrix} Q_1 A_1^h A_1^{-h} Q_1^{-1} A_1^{-1} A_1 \hat{S}_{\omega,t} \\ O \end{pmatrix} \\ &= \begin{pmatrix} \hat{S}_{\omega,t} \\ O \end{pmatrix} \end{aligned} \quad (35)$$

と計算され, Q_1 に依存せず, 真の音源信号である式 (30) に一致することがわかる。

上記の性質より, 時間周波数毎に見ると劣決定の条件下で解を決めているのは $Q_{\omega,t}^{-2}$ の対角成分であり, 性質 2 から, 定性的には時間周波数マスクのように働くことがわかる。ただし, 1) 通常の時間周波数マスクは 0~1 の値をとりその絶対的な値が信号を通過させる割合の意味をもつが, 性質 1 から $Q_{\omega,t}^{-2}$ の対角成分は相対的な値のみに意味がある, 2) 性質 4 より, 必要とされる音源信号のスパース性は通常の時間周波数マスクより緩和されており, 観測信号数と等しい混合までは, 正しい解を与える, 3) 性質 3 より, 対角成分が等しい場合はノルム最小解を与える, 等の特徴があることがわかる。

本稿では, 周期信号の定義である $f(t) = f(t+T)$ ヘフィッティングした残差をガウス分布でモデル化することにより, 式 (10) の形を導いたが, 本節で議論した性質に基づき, 基本周期 $T_{j,t}$ に対して $Q_{\omega,t}^{-2}$ を別な形で設計することも考えられる。

5. 評価実験

5.1 実験条件

本稿で議論した調波成分分析によるブラインド音源分離の基礎的な確認と分離性能の評価のために実験を行った。音楽信号に多く含まれる, 高調波同士が重なりあう音源信号として, 各時間フレームにおいて, 2 つが常に 1 オクターブ差で, もう 1 つがその間を半音ずつ変化するような信号を MIDI で作成し, これを WAV フォーマットに変換した。サンプリング周波数は 16kHz である。音色はすべてバイオリンとし, また音源信号の基本周波数は互いに入れ替わるような設定とした。つまり, 音源信号の音色や時間方向の連続性は音源分離の手がかりとして用いることができず, 混合行列の時不変性と調波性のみを手が

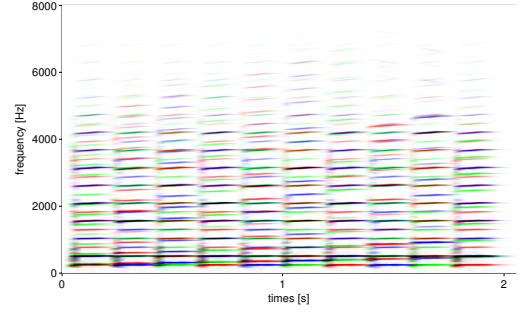


図 1 音源信号のスペクトログラム。赤, 緑, 青がそれぞれ, 音源 1, 2, 3 に対応する

かりに分離しなければならない例となっている。スペクトログラムを図 1 に示す。ところどころで高調波同士の重なりが確認できる。また, 本稿で論じた調波成分分析は, 畳込み混合も扱える枠組みになっているが, 本実験では音量バランスを変えて remix されたステレオ信号を想定して瞬時混合とし, 音量バランスのみ異なる混合行列を与えた。時間周波数分解は, フレーム長 1024 点, シフト 512 点で, Hamming 窓関数を用いた短時間 Fourier 変換により行った。

5.2 結果と考察

まず, 混合行列の推定結果を表 1 に, ピッチ周期の推定結果を図 2 に示す。混合行列の推定はほぼ正しく行われておるが, ピッチ周期の推定はところどころ誤りを含み, 特に半ピッチの誤りが多くみられる。本稿での定式化では, 楕円ノッチフィルタの出力パワーが小さくなることのみで調波性をはかっているため, このような半ピッチの誤りは避けられないが, 今後は分離された音源信号の高調波のパワーの分布も考慮して調波信号の事前分布をモデル化することにより改善できる可能性があると考えている。

次に, SN 比改善値を表 2 に示す。比較のために, 理想的なバイナリマスクングを用いた結果も示す。ここで理想的とは, 各時間周波数成分における音源信号のパワーを既知としてバイナリマスクを設計したという意味であり, バイナリマスクングの性能上限を示しているが, 実際には各時間周波数成分における音源信号のパワーは未知であり, 観測信号間の強度比や時間差のクラスタリングによって設計されたバイナリマスクでは, 分離性能はここで示す上限値よりも低下することに注意する。

今回の実験では, 基本周期 (基本周波数) の推定に半ピッチ誤りが多く含まれていたが分離結果の SN 比としては平均して 10[dB] 程度の改善がみられており, これはほぼ, 理想的なバイナリマスクングに匹敵する。基本周波数の推定結果は, 音源信号の推定式である式 (17) の中の $Q_{\omega,t}^{-2}$ という行列の対角成分を決定し, これは 4 節での議論の通り, 定性的には時間周波数マスクのように働く。正しい基本周波数 ω_0 に対し半ピッチの誤りは, 実際には含まれていない $(2n+1)\omega_0/2$ のような周波数も通過させるように働くが, 性質 4 より, 通過と推定された音源が 2 つまでであれば, 混合行列 A の部分的な逆行列により, 正しく音源信号が推定されるため, このような半ピッチ誤りが大きな弊害になっていないものと考えられる。例えば音源信号

表 1 混合行列 A の推定結果

要素	(1,1)	(2,1)	(1,2)	(2,2)	(1,3)	(2,3)
推定値	0.989	0.150	0.765	0.644	0.258	0.966
真値	0.985	0.174	0.766	0.643	0.342	0.940

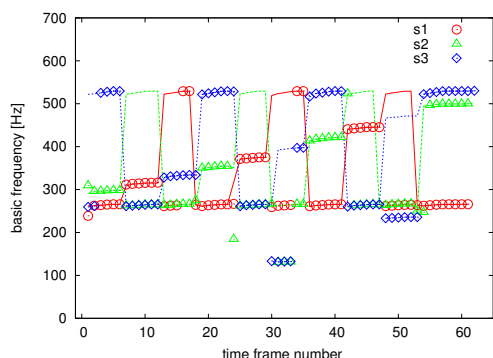


図 2 基本周波数の推定結果。点は推定値，線は真値を表し，音源を色で区別している。

表 2 観測信号・分離信号の SN 比 [dB]

	音源 1	音源 2	音源 3
観測信号 (ch1)	1.6	-2.3	-9.1
観測信号 (ch2)	-10.0	-3.2	3.0
分離信号	13.2	8.4	13.5
理想バイナリマスク (ch1)	11.1	11.4	2.9
理想バイナリマスク (ch2)	3.1	9.5	12.1

2 に対して，分離信号のスペクトログラムと音源信号自体のスペクトログラムを 3, 4 に示す。10 フレーム目付近や 25 から 30 フレーム付近、また 45 フレーム付近において，音源信号 s2 の基本周波数推定結果は真値の半分程度の値に推定されているが，スペクトログラムの推定結果においては推定された基本周波数の奇数倍の周波数成分（つまり，真の調波成分ではない周波数成分）はあまり含まれていないことがわかる。

6. おわりに

本稿では，音源信号の調波性をスペクトログラムの二次形式で表されるような事前分布に基づき，調波性の最大化に基づく劣決定ブラインド音源分離の枠組みについて論じ，具体的なパラメータ推定の更新式を導出し，基礎的な実験により分離性能を評価した。今後は，勾配法で推定している混合行列 A の推定的高速化，調波信号の事前分布を与える確率モデルの改良などを検討しつつ，畳込み混合や実際の音楽信号への適用実験を行っていく予定である。

文 献

- [1] http://www.music-ir.org/mirex/2007/index.php/Main_Page
- [2] <http://sassec.gforge.inria.fr/>
- [3] E. Vincent *et al.*, “First Stereo Audio Source Separation Evaluation Campaign: Data, Algorithms and Results,” Proc. Int. Conf. on Independent Component Analysis and Blind Source Separation (ICA), Springer, pp. 552-559, 2007.
- [4] Y. Haraguchi, S. Miyabe, H. Saruwatari, K. Shikano, T. Nomura, “Evaluation of Sound-Localization Control of Audio Objects Based on Blind Source Separation,” 電子情報通信学

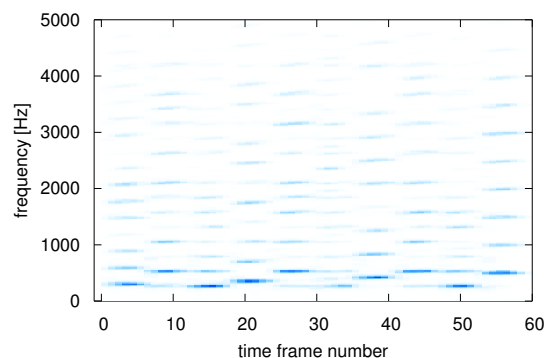


図 3 音源信号 2 の推定されたスペクトログラム

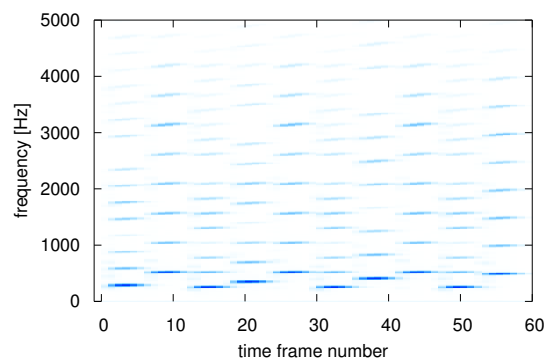


図 4 音源信号 s2 の真のスペクトログラム

会技術研究報告, EA2007-56, vol.107, no.240, pp. 7-12, Sep., 2007.

- [5] A. Hyvärinen *et al.*, “Independent Component Analysis,” John Wiley, New York, 2001.
- [6] O. Yilmaz and S. Rickard, “Blind Separation of Speech Mixtures via Time-Frequency Masking,” IEEE Transaction on Signal Processing, Vol. 52, No. 7, pp 1830-1847, 2004.
- [7] S. Araki, H. Sawada, R. Mukai, and S. Makino, “Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors,” Signal Processing, vol. 87, pp. 1833-1847, Feb. 2007.
- [8] Y. Izumi, N. Ono, and S. Sagayama, “Sparseness-based 2ch BSS using the EM algorithm in reverberant environment,” Proc. WASPAA, pp.147-150, Oct., 2007.
- [9] 西, 安部, 安藤, “聴覚情景分析のための多重ピッチ追跡と調波分離アルゴリズム,” 計測自動制御学会論文集, vol.34, no.6, pp.483-490, 1998.
- [10] 安部 素嗣, 安藤 繁, “共有 FM-AM の時間周波数統合に基づく聴覚情景解析 (I) — Lagrange 微分特微量とその周波数軸統合 —,” 電子情報通信学会論文誌, vol.J83-D-II, no.2, pp.458-467, 2000.
- [11] H. Kameoka, T. Nishimoto, and S. Sagayama, “A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering,” IEEE Trans. on Audio, Speech and Language Processing, Vol. 15, No. 3, pp.982-994, Mar., 2007.
- [12] H. Kameoka, N. Ono, and S. Sagayama, “Auxiliary Function Approach to Parameter Estimation of Constrained Sinusoidal Model,” Proc. ICASSP2008, pp.29-32, Apr., 2008.
- [13] M. Nakashizuka, “A Sparse Decomposition Method for Periodic Signal Mixtures,” IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. E91-A(3), pp. 791-800, 2008.