

Title	複数音源中からの目的楽器音の選択的分離抽出に関する研究
Author(s)	窪, 正晃
Citation	
Issue Date	2003-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1676
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

修 士 論 文

複数音源中からの
目的楽器音の選択的分離抽出に関する研究

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

窪 正晃

2003年3月

修 士 論 文

複数音源中からの
目的楽器音の選択的分離抽出に関する研究

指導教官 赤木 正人 教授

審査委員主査 赤木 正人 教授
審査委員 小谷 一孔 助教授
審査委員 下平 博 助教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

110040 窪 正晃

提出年月: 2003 年 2 月

概要

本論文では、楽器音の音響的特徴を知識として利用し、複数の楽器音が混在する音の中から目的音を選択的に分離抽出するモデルを提案する。これは、人が聴きたい音に“聞耳をたてる”とき、既知の目的音情報を利用しているという仮定に基づいたものである。提案するモデルは、楽器音の音響的特徴を反映したテンプレートを利用して目的音の選択を行うトップダウン処理と、目的音とその他の音源の周波数成分を分離するボトムアップ処理を融合したものである。

分離モデルを評価するために、(1) 目的音に雑音が付加された状況、(2) 二楽器音が混合した状況、(3) 四種類の楽器音が混合した状況から、それぞれの目的音を選択的に分離抽出するシミュレーションを行った。この結果、提案するモデルが混合音中から目的音を選択し、高い精度で分離抽出できることが示された。

目次

第1章	序論	1
1.1	はじめに	1
1.2	研究の背景	2
1.2.1	計算論的聴覚情景解析の研究	2
1.2.2	数理工学的な信号分離の研究	4
1.3	本論文の目的	4
1.4	本論文の構成	5
第2章	選択的分離抽出問題の解法	6
2.1	問題設定	6
2.1.1	仮定	6
2.2	選択的分離抽出モデル	7
2.3	モデルによる処理の概要	9
2.3.1	知識を利用したトップダウン処理	9
2.3.2	二波形分離モデルによるボトムアップ処理	11
第3章	分離抽出モデルの実装	15
3.1	周波数分析部	16
3.2	基本周波数推定部	16
3.3	特徴量テンプレート生成部	18
3.4	目的音選択部	23
3.5	波形分離部	25
3.6	グルーピング部	26
第4章	分離モデルの評価	27
4.1	シミュレーション条件	27
4.1.1	評価尺度	27
4.1.2	比較対象	28
4.2	楽器単音の分離	28
4.2.1	雑音中の目的音分離抽出	29
4.2.2	二楽器混合音からの分離抽出	32

4.2.3	複数楽器混合音からの選択的分離抽出	35
4.3	楽器演奏音の分離抽出	41
第5章	結論	44
5.1	まとめ	44
5.2	今後の展望	45

目次

1.1	聴きたい音を聞き分ける : カクテルパーティ効果	1
2.1	選択的分離抽出モデルの構成	8
2.2	モデルによる信号処理の流れ	10
3.1	モデルの構成	15
3.2	分析合成系の評価	17
3.3	瞬時振幅 $S_k(t)$ と $S_k(t)$ の自己相関関数	19
3.4	推定された基本周波数候補	20
3.5	フルート、ピアノ、ヴァイオリン分析例	22
3.6	フルート、ピアノ、ヴァイオリンの標準的なテンプレート形状	23
3.7	目的音選択部での信号処理の流れ	24
4.1	分離例 (フルート (A4)+白色雑音)	30
4.2	雑音中からの目的音分離抽出結果	31
4.3	分離例 (フルート (A4)+ヴァイオリン (C3))	33
4.4	二楽器混合音からの目的音の分離抽出結果	34
4.5	四種類の楽器音からの目的音分離例 (フルート)	36
4.6	四種類の楽器混合音からの分離精度 (フルート)	37
4.7	四種類の楽器混合音からの分離精度 (ピアノ)	38
4.8	四種類の楽器混合音からの分離精度 (ホルン)	39
4.9	四種類の楽器混合音からの分離精度 (ヴァイオリン)	40
4.10	楽器演奏音の分離抽出例	43

表 目 次

2.1	四つの発見的規則と制約条件の関係	12
3.1	制約条件とその定式化	25
4.1	楽器演奏音の混合に用いた楽器とその旋律	41

第1章 序論

1.1 はじめに

我々は、混在する多くの音の中から聴きたい音を選択し、聴き分けることができる。この能力はカクテルパーティ効果と呼ばれ、人間の聴覚システムが持つ優れた能力の一つとして知られている。この名前は、混雑したカクテルパーティ会場において、人々の会話、グラスの触れ合う音、会場に流れるBGM、といった様々な音が渦まいた状況においても、話し相手の声のみを聴き取り会話ができるということに由来する。ただし、この能力の対象は人間の音声に限らず、様々な音に対してその効果を発揮する。例えば、人々が行き交う雑踏の中にあっても硬貨が地面に落下するときには、周囲の誰もがその特徴的な音を聴き取る事ができ、その方向に振り返るといった反応が予想できる。また、楽器アンサンブルやオーケストラ演奏など音楽を楽しむ時には、いくつもの音の重なり合いを味わうことはもちろんだが、アンサンブル中のある楽器のソロパートに注目して聴き取り、その旋律の美しさを味わうことも音楽観賞の重要な楽しみである。

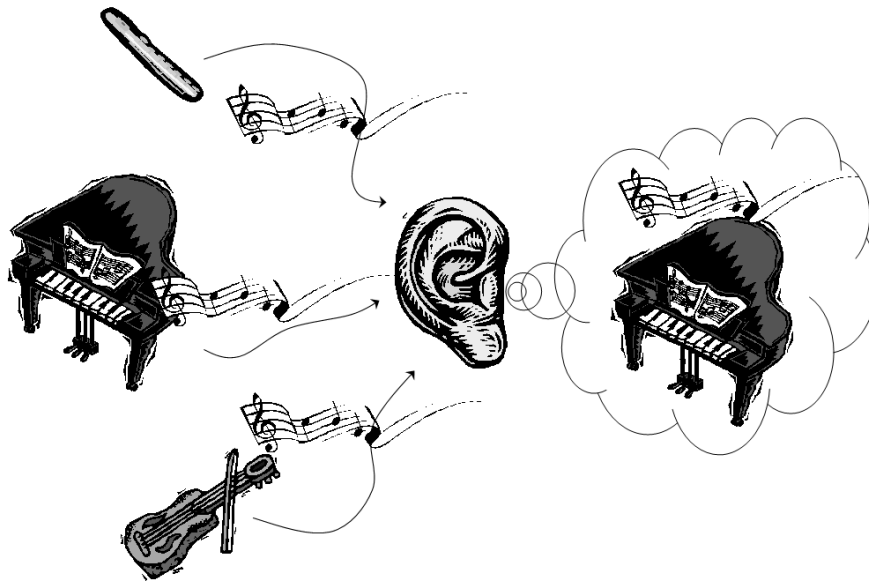


図 1.1: 聴きたい音を聞き分ける : カクテルパーティ効果

このように、人間は音声や音楽といった外界に生じる様々な音の中からカクテルパーティ効果などの聴覚能力により聴きたい音を選択して聴き、それぞれの事象を把握し知覚することができる。人間が聴覚機能により自らの周囲を取り巻く環境を把握する能力は聴覚の情景解析 (ASA: Auditory Scene Analysis) と呼ばれ、Bregman の著書 [1] によって広く知られるようになった。この名前は視覚研究における情景解析に由来するものである。

近年、人間の持つ聴覚情景解析の優れた機能を計算論的アプローチから解明し、計算モデルとして実現することを目的に様々な研究が行われている。カクテルパーティ効果に代表される聴覚の優れた情報処理機能が、計算モデルとして実現できれば様々な分野への応用が期待できる。例えば、工学的な応用として、自律ロボットの耳として人間と同等の聴覚機能を計算機上に実現する場合や、雑音に頑健な音声認識システムのフロントエンドへの応用が考えられる。また、音楽音響信号を対象に望みの楽器音を分離抽出することが可能となれば、自動採譜システムやコンピュータミュージックなどへの応用が期待できる。

しかし、現状では音源が複数存在しそれらが混合した状況から、目的音を選択的に分離抽出するという問題を完全に実現するには至っていない。その問題点としては、(1) 混合音中の複数の音源はすべて分離抽出の対象となる可能性があるため、それらの中から目的音を選択し、分離抽出する必要があること、(2) 目的音の成分と他の音源の成分が周波数領域で重複する場合、重複周波数成分から目的音の周波数成分のみを分離する必要があること、があげられる。このように、複数の音源から目的音を選択的に分離抽出するという音源分離問題を解くためには上記の二つの問題を解決する必要がある。

では、我々が聴きたい音を聴く場合にはどのように上記の問題を解いているのだろうか。我々が、たくさんの音の中から聴きたい音に注意を向け選択的に聴こうとする行為を表現するものとして、「聞耳をたてる」という言葉がある。本論文では、我々が聴きたい音を聴くときに行う「聞耳をたてる」という行為をヒントに、この「聞耳」模擬する一手法を提案する。提案法では、我々が聴きたい音に「聞耳をたてる」際には目的音に関する何らかの情報を利用しているのではないかという仮定に基づき、目的音の音響的特徴を知識として利用する選択的分離抽出モデルを提案する。ここでは、複数の音源を含む音響信号として音楽音響信号を対象とし、複数の楽器音が混合した信号中から目的の楽器音を選択的に分離抽出する音源分離モデルの構築を試みる。

1.2 研究の背景

1.2.1 計算論的聴覚情景解析の研究

計算論的聴覚情景解析 (CASA: Computational Auditory Scene Analysis) という研究分野では、人間の聴覚に備わる優れた情報処理機能である聴覚情景解析の能力を計算モデルとして実現することを目的に多くの研究が行われている。これらの研究では、カクテルパーティ効果を音源分離問題として計算機上に実装する試みが行われている。

しかし、前節で述べたように複数音源からの目的音の選択的分離抽出を実現するために

は以下のことを考えなければならない。

- (1) 複数の音源中から目的音を選択しなければならない
- (2) 目的音成分と他音源成分との重複成分を分離しなければならない

このような課題を含む選択的分離抽出問題に対してこれまでに提案されている音源分離モデルは、大きく分けてAI的アプローチによるトップダウン処理に基づくものと、信号処理的アプローチによるボトムアップ処理に基づくモデルが提案されている。

トップダウン処理に基づくモデルには、心理音響学的なグルーピング規則などを利用するモデル [8] や、マルチエージェントシステムを用いたモデル [9] などがある。Ellis のモデルでは、一つの音事象に関する調波性、共通の立上り、近接、連続性という基本的な規則を用いて、混合音中に存在するそれぞれの音事象に属する周波数成分をグルーピングすることで音源分離を実現している。また、中谷らのモデルでは混合音中のそれぞれの音事象の基本波成分を追跡するマルチエージェントを用いて連続的な音響ストリームを抽出し、その倍音関係となる周波数成分を抽出することにより音源分離を実現している。以上のようなモデルでは複数の音源からそれぞれの音事象を抽出することを可能としている。しかし、音響的な特徴として振幅（もしくはパワー）スペクトルのみを利用しているため、これらのモデルでは、二つの信号成分が同一周波数領域にある場合、完全に分離できているとは言い難い。

一方、ボトムアップ処理に基づくモデルとして、ベイジアンネットワークを用いたモデル [6] や、二波形分離モデル [3] などが報告されている。

柏野らのモデルでは、ベイジアンネットワークを用いた仮説ネットワークによるボトムアップ処理とトップダウン処理の情報統合を試みている。このモデルではまず、ボトムアップ処理によって得られた周波数成分から調波性、立上りの同期などを利用し混合音中に含まれる単音仮説を生成する。さらに、楽器音の音色、周波数成分の構成といった知識を用いて仮説ネットワークにおいて事後確立最大となる仮説を決定し、MIDI (Musical Instrument Digital Interface) データとして記号表現による出力を得ている。しかし、このモデルにおいても重複周波数成分の分離は行われておらず、仮説生成や音源同定処理の性能悪化の原因となっている。また、モデルの出力として MIDI データによる記号表現以外に、抽出された周波数成分を再合成し音響信号を得ることも可能であるが、この場合にも重複周波数成分の影響により完全な分離ができているとは言い難い。

それに対し、二波形分離モデル [3, 4, 5] では、振幅 (パワー) スペクトルに加えて位相スペクトルを用いることによって、重複周波数成分の分離を実現している。二波形分離モデルでは、目的音成分と雑音成分が同じ周波数領域に存在する状況を含む音源分離問題を二波形分離問題として定式化し、さらに Bregman が提唱した四つの発見的規則 [2] を制約条件として利用することで、目的音と雑音の周波数成分の分離を可能にしている。しかし、二つの音源がどちらでも分離対象となりえるような問題設定ではないことから、上記の問題 (1) には対応しきれていない。

以上のようにそれぞれのモデルは、二つの問題のうちのいずれか一方に対応するのみであり、両者の問題を同時に解決できるモデルはまだないのが現状である。

1.2.2 数理工学的な信号分離の研究

上記のような人間の聴覚機能の解明を目的とした研究以外にもカクテルパーティ効果のような音源分離問題を数理的定式化し解くことを目的とした研究が数多く行われている。

その例として、ここでは独立成分分析 (ICA:Independent Component Analysis) という情報理論に基づいた音源分離手法 [15] を取りあげる。この手法は線形作用によって混合されたいくつかの独立成分の復元が目的であり、一般的には以下のように定式化される。我々がマイクなどのセンサーによって得る混合信号 $\mathbf{X}(t) = [x_1(t), \dots, x_m(t)]^T$ は、 n 個の信号源 $\mathbf{S}(t) = [s_1(t), \dots, s_n(t)]^T$ から、ある線形メディアを通して伝搬してきたものとすると、次のような線形モデル

$$\mathbf{X}(t) = \mathbf{A}\mathbf{s}(t) \quad t = 1, 2, \dots \quad (1.1)$$

で表現できる。このモデル化において混合行列 \mathbf{A} の逆行列を推定し観測信号 $x(t)$ から原信号 $s(t)$ を引き出すことがこの音源分離という逆問題を解く方針である。この手法では、それぞれの音源が独立であることが制約条件であり、また、観測するセンサーの数は信号源の数と同じか、またはそれより多くなければならないという強い制約が必要である。つまり、モノラルの音響信号からの二つの音源を分離することには対応できないということである。

また、本論文の目指す選択的分離抽出問題の二つの問題点に対応づけた場合を考えると、この方法では上記の制約条件を克服することができれば、(2) の問題である重複周波数成分の分離が実現する可能性がある。しかしながら、(1) の問題である目的音の選択という課題について、この手法では目的音がどのような音源であるかを考慮しないブラインド音源分離であるため、目的音を選択的に分離するという手法ではない事がわかる。以上より、聴きたい音を聴くという行為の模倣を目指す本論文の目的とは外れるものの、工学的な応用を考えた場合には非常に有用な手法であるといえる。

1.3 本論文の目的

前節に示したように、これまでに提案されているモノラル音響信号を対象とした音源分離モデルでは、複数の音源が混合した状況から、目的音のみを選択的に分離抽出するという問題を完全に実現するには至っていない。

本論文では、選択的分離抽出を実現するために必要な二つの問題を解決する方法として、目的音の音響的特徴を利用する手法を提案する。提案法では、我々が聴きたい音を聴

く際には目的音に関する何らかの情報を利用することで聞き分けを行っているのではないかと考え、分離抽出の手がかりとしている。

実際に提案法を実装し選択的分離抽出を実現するモデルとして、音源間の重複周波数成分の分離が可能なボトムアップ処理（二波形分離モデル）と、目的楽器音のパワーやスペクトル形状などの音響的情報を知識として利用し抽出を行うトップダウン処理を融合した音源分離モデルを提案する。

1.4 本論文の構成

本論文は5章で構成される。第1章では、本論文で扱う音源分離問題に対する研究の背景と問題点を明らかにし、本論文の特色と目的を示す。

第2章では、提案する選択的分離抽出モデルについて述べる。はじめに、本論文で対象とする選択的分離抽出問題の問題設定を行う。その上で、聴きたい音を聴く際の経験的事実から得られる仮定と、その仮定を利用した選択的分離抽出モデルによる問題の解法を述べる。

第3章では、第2章で提案した選択的分離抽出モデルを実装し、その構成を示す。また、モデルを構成する周波数分析部、基本周波数推定部、特徴量テンプレート生成部、目的音選択部、波形分離部の詳細について述べる。

第4章では、実装した選択的分離抽出モデルによる分離シミュレーションを行いモデルの有効性を検証する。

第5章では、本論文で得られた結果をまとめ、今後の展望を述べる。

第2章 選択的分離抽出問題の解法

この章では、本論文で扱う選択的分離抽出問題の問題設定を行う。その上で、問題を解くために設けた仮定とその基本的コンセプトを示し、仮定を利用した選択的分離抽出モデルについて述べる。

2.1 問題設定

我々人間は様々な音が飛び交う日常のなかで、カクテルパーティ効果などの聴覚の機能を用いることで難なく目的音を選択し聴き分けている。例えば、我々がピアノとバイオリンによるアンサンブルを聞いた場合に、特別な訓練を積まなくてもピアノとバイオリンそれぞれの音を容易に聴き分ける事ができる。カクテルパーティ効果のメカニズムはまだ完全に解明されたわけではないが、到来方向の違い、音源のピッチの違い、音色の違いといったことが関係していると考えられている [25]。

本論文の目的とする選択的分離抽出とは、このカクテルパーティ効果を模擬する一例であり、混在する多くの音の中から目的音の音響信号を分離する音源分離問題である。

第1章でも述べたように、このような選択的分離抽出問題には二つの問題点が存在する。一つ目は、たくさんの音の中からどの音を取り出すかを選択するという問題であり、これは聴きたい音を聴くという人間の注意を表現する課題である。二つ目は、重なり合う音の中からどのようにして目的音を分離抽出するかという問題である。これらの問題を計算機上で実現する事を考えた場合には、複数存在する音源から目的音を選択するための音源同定処理と、混合された信号中から目的音を分離抽出する音源分離処理が必須となる。しかし、これまでの手法ではそれぞれの音が重なり合うことにより、同一周波数成分を持つ場合には高精度の処理が困難であった。

では、我々が聴きたい音を聴く場合にはどのような方法を用いているのだろうか。本論文では、聴きたい音を聴く際の経験的な事実から以下のような仮定を設け、分離抽出の手がかりとして利用する事とした。

2.1.1 仮定

我々が日常生活の中で混合音の中から目的音を選択的に聴こうとするとき、その音をよく知っている場合や、直前に一度聞いておいた場合、比較的容易に目的音を聴き取ること

ができることがある。本論文では、この経験的な事実から、“人が聴きたい音を聴き取る
とき、目的音に関する何らかの事前情報を利用しているのではないか”と考え、聴きたい
音を聴く“聞耳”を表現するために次のような二つの仮定を行った。

1. 混合音中のどこかに目的音が必ず存在する。
2. 目的音の音響的特徴を知識として持っている。

この二つの仮定は、目的音がどのような音かをあらかじめ知っている場合に、混合音中
の目的音に「聞耳」をたてて聴き分けようとする状況を表現するものである。音楽を聞く
場合に置き換えると、ピアノという楽器の音がどのような特徴を持つ音かを知っている場
合に、ピアノとバイオリンのアンサンブル演奏中のピアノの音のみを聴き分けるような状
況を仮定している。

知識として与える音響的特徴とは、各楽器音の振幅スペクトルや各調波成分の基本波
に対するパワー比などである。ただし、この音響的特徴は各楽器音の一般的な特徴であっ
て、混合音中に含まれる目的楽器音と完全に一致する必要はない。また、各楽器音の一般
的な特徴を用いる事で同種楽器の個体差や音高による違いを吸収するという狙いも含ま
れている。提案法では、各楽器音のおおよその特徴量を利用して混合音中に含まれる本来
の目的音を推定し分離するため、得られる分離音は知識として持っている特徴量そのもの
とはならない。

知識として利用する各楽器の具体的な特徴量については3.3節で述べる。

この仮定に基づき目的音の音響的特徴を知識として利用する選択的分離抽出モデルを
提案する。

2.2 選択的分離抽出モデル

前節で述べたように、我々が普段難なく実現している選択的分離抽出を計算機上で実現
する事は非常に困難な問題である。この問題に対して本論文では、目的音に関する知識を
手がかりとした選択的分離抽出モデルにより目的音の選択と分離の両者を実現する。

実装した選択的分離抽出モデルの構成を図2.1に示す。提案するモデルは(1)周波数
分析部：FiterBank、(2)基本周波数推定部：Fnote Estimation、(3)特徴量テンプレ
ート生成部：Template ganerator、(4)目的音選択部：Event Detection、(5)波形分離
部：Separation block、(6)グルーピング部：Grouping blockの六つの処理ブロックで
構成される。モデルの入力は、混合信号 $f(t)$ と目的楽器名 (Symbol) である。また、仮定
1,2により楽器音の音響的特徴を知識源として与えており、これを目的音に関する情報と
して利用する。

各処理ブロックの実装など詳細な点は第三章で述べるが、選択的分離抽出モデルによる
大まかな処理の流れと、各処理ブロックの役割は以下のようになる。

1. 周波数分析部 (FiterBank)：観測された混合音の周波数分析（瞬時振幅を瞬時位相の

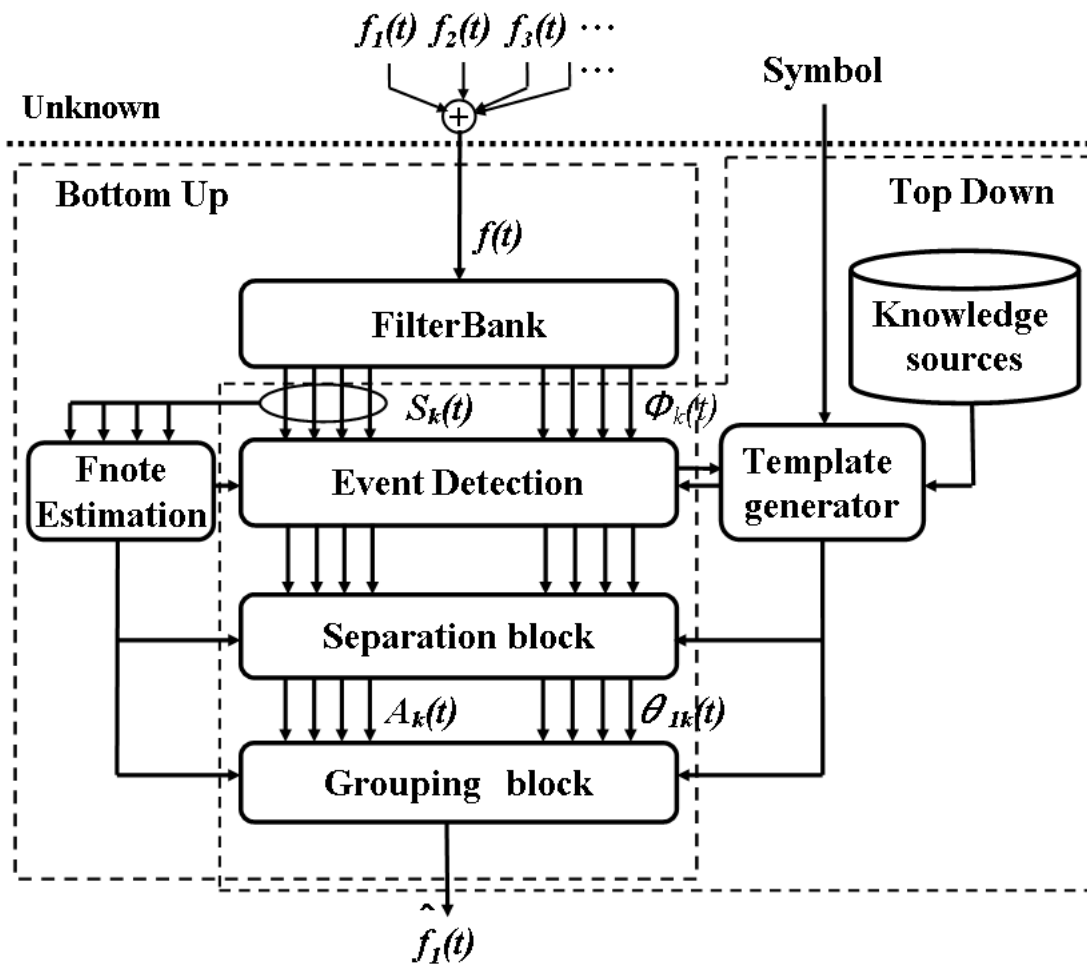


図 2.1: 選択的分離抽出モデルの構成

- 抽出)を行う。
2. 特徴量テンプレート生成部 (Template generator) : Symbol 入力に従い知識源より目的音に対応したテンプレートを生成する。
 3. 基本周波数推定部 (Fnote Estimation) : 混合音中に含まれる音源の基本周波数を推定する。
 4. 目的音選択部 (Event Detection) : 混合音の周波数成分、基本周波数、特徴量テンプレートを用い目的音を選択する。
 5. 波形分離部 (Separation block) : 混合信号の周波数成分から目的音のものと他の音源のものを分離する。
 6. グルーピング部 (Grouping block) : 分離された目的音の周波数成分をまとめ、それらを再合成し目的音の音響信号を出力する。

また、モデルによる信号処理は図 2.2 に示すような流れで行われる。

2.3 モデルによる処理の概要

提案するモデルは図中の破線で示すように、混合音中の目的音の位置等を選択するトップダウン処理と、目的音とその他の音源の周波数成分を分離するボトムアップ処理を融合したものである。ボトムアップ処理部には目的音とその他の音源が周波数領域で重複する場合でも分離が可能な二波形分離モデル [3] を利用している。しかし、このモデルは入力を二波形に限定しているため、本論文で取り扱うような複数の音源波形が混合した信号には直接利用できない。

そこで提案するモデルでは、まず仮定より得られた目的音の音響的特徴を利用して混合音中の目的音を見つけ出す。これにより混合音は選択された目的音波形と、それ以外すべての音源波形の二波形とみなすことができ、二波形分離問題 [5] として解くことが可能となる。

ここでは、提案するモデルによる分離抽出処理の概要をトップダウン処理とボトムアップ処理のそれぞれについて述べる。

2.3.1 知識を利用したトップダウン処理

はじめに、目的音の音響的特徴を利用し目的音の選択を行うトップダウン処理の流れを述べる。混合音中からの目的音成分の探索は以下のような手順で行われる。まず、モデルの知識源に蓄えられた各楽器によって異なる音響的特徴量を利用して、目的とする楽器の種類に応じた特徴量テンプレートを生成する (特徴量テンプレート生成部 : Template Generator)。テンプレートは、楽器音の持つ調波成分の振幅包絡形状と、各調波成分の基本波に対するパワー比を反映した時間周波数平面上での形状で表現されている。

このテンプレートを利用し、混合音中から目的音成分を選択することとなる (目的音選

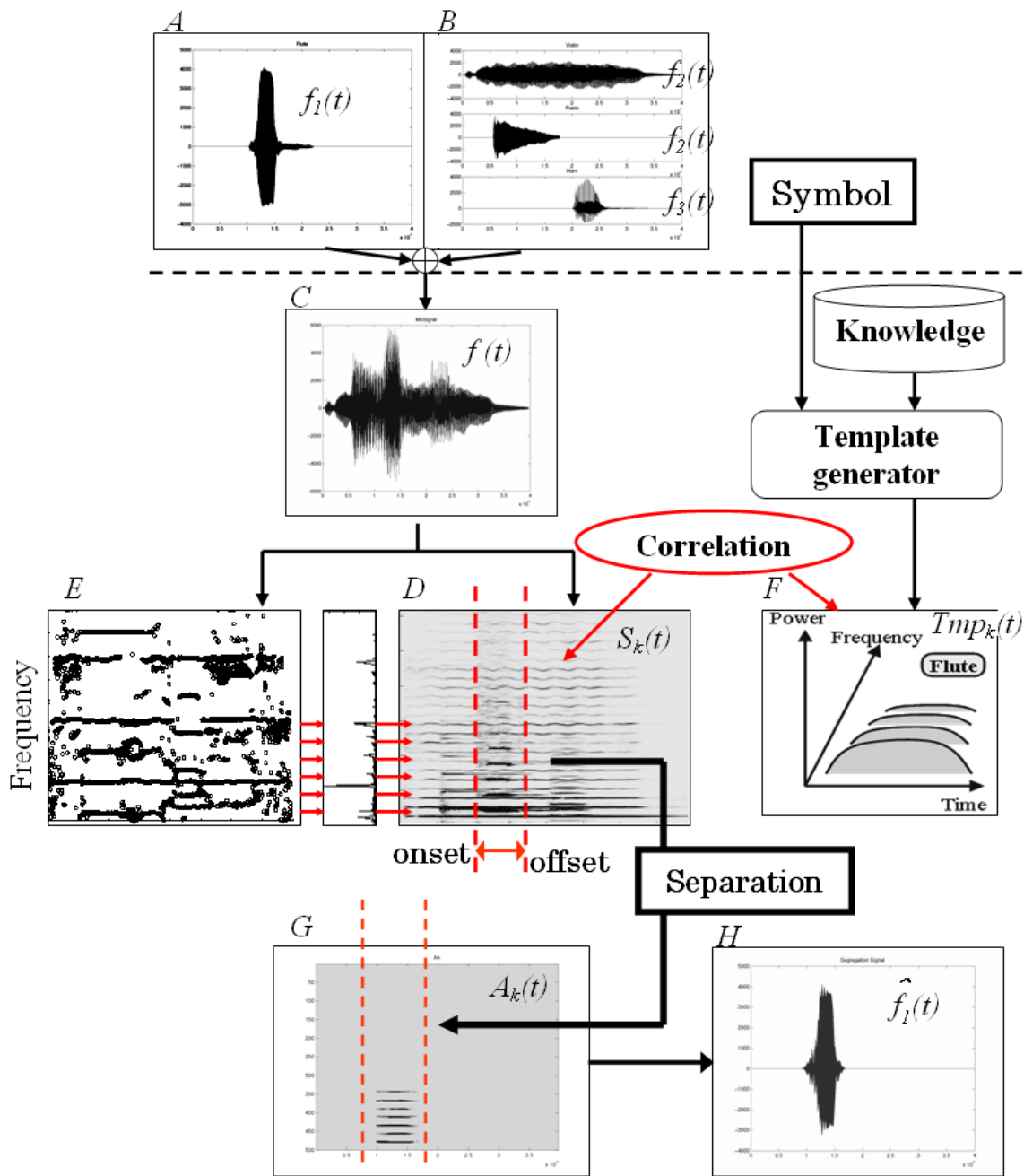


図 2.2: モデルによる信号処理の流れ

扱部：Event Detection)。まず、周波数分析部で得られた混合音の周波数成分と基本周波数推定部で得られる基本周波数の二つの情報から、混合音中に含まれる音源成分を抽出し目的音候補とする。次に、これら候補の中で最も目的音のテンプレートと相関が高いものを目的音成分として選択する。この時テンプレートは、比較対象となる目的音候補の持続時間に応じてその形状を変化させている。

以上の処理によって Symbol 入力によって示された目的音が選択され、その目的音成分の時間領域での位置と基本周波数が特定できる。

しかし、ここまでのトップダウン処理のみで得られた目的音成分には従来のモデルで問題となっていた他音源の重複周波数成分が存在している。そこで、本モデルでは次のようなボトムアップ処理によりその分離を行う。

2.3.2 二波形分離モデルによるボトムアップ処理

以上のトップダウン処理により、混合音中から目的音の位置とその基本周波数を得ることができた。これにより、混合音は選択された目的音波形と、それ以外すべての音源波形の和とみなすことができ、ここで以下に示す二波形分離問題の定式化 [3] に従って解くことができる。

はじめに、目的音の $f_1(t)$ と、それ以外のすべての音源の混合音 $f_2(t)$ が加算され、混合音 $f(t) = f_1(t) + f_2(t)$ のみを受音できるものとする。これを K 個の分析フィルタ群により周波数分解したとき、 k 番目の分析フィルタを通過した $f_1(t)$ と $f_2(t)$ の周波数成分をそれぞれ

$$f_1(t) : X_{1,k}(t) = A_k(t) \exp(j\omega_k t + j\theta_{1k}(t)) \quad (2.1)$$

$$f_2(t) : X_{2,k}(t) = B_k(t) \exp(j\omega_k t + j\theta_{2k}(t)) \quad (2.2)$$

と仮定すれば、 $f(t)$ の通過成分 $X_k(t)$ は

$$X_k(t) = X_{1,k}(t) + X_{2,k}(t) \quad (2.3)$$

$$= S_k(t) \exp(j\omega_k t + j\phi_k(t)) \quad (2.4)$$

と表される。ただし、 ω_k は分析フィルタの中心角周波数、 $A_k(t)$ 、 $B_k(t)$ 、 $S_k(t)$ は瞬時振幅、 $\phi_k(t)$ は瞬時出力位相、 $\theta_{1k}(t)$ と $\theta_{2k}(t)$ は瞬時入力位相である。ここで瞬時振幅 $S_k(t)$ と出力位相 $\phi_k(t)$ は、それぞれ

$$S_k(t) = \sqrt{A_k^2(t) + 2A_k(t)B_k(t) \cos \theta_k(t) + B_k^2(t)} \quad (2.5)$$

$$\phi_k(t) = \arctan \left(\frac{A_k(t) \sin \theta_{1k}(t) + B_k(t) \sin \theta_{2k}(t)}{A_k(t) \cos \theta_{1k}(t) + B_k(t) \cos \theta_{2k}(t)} \right) \quad (2.6)$$

表 2.1: 四つの発見的規則と制約条件の関係

発見的規則 (Bregman,1993)	制約条件 (Unoki and Akagi,1999)
(i) 関連のない音が一緒に始まったり 終わったりすることはない	立上り, 立下りの同期
(ii) 変化は急激には起こらない	漸近的变化
(a) 一つの音の属性はゆっくりとなめらかに 変化する傾向がある	多項式近似+なめらかさ
(b) 同じ音源から生じる音の一連の属性は ゆっくりとなめらかに変化する傾向にある	多項式近似+なめらかさ
(iii) 物が繰り返し振動するときには、 共通の基本周波数の 整数倍の音響成分が発生する	調波関係
(iv) 一つの音響事象に生じる多くの変化は、 その音を構成する各成分に 同じような影響を与える	振幅包絡 $A_k(t)$ 間の相関

で求められるため、瞬時振幅 $A_k(t)$ と $B_k(t)$ は、

$$A_k(t) = \frac{S_k(t) \sin(\theta_{2k}(t) - \phi_k(t))}{\sin(\theta_{2k}(t) - \theta_{1k}(t))} \quad (2.7)$$

$$B_k(t) = \frac{S_k(t) \sin(\phi_k(t) - \theta_{1k}(t))}{\sin(\theta_{2k}(t) - \theta_{1k}(t))} \quad (2.8)$$

として解くことができる。ただし $\theta_k(t) = \theta_{1k}(t) - \theta_{2k}(t)$ であり、 $\theta_k(t) \neq n\pi, n \in Z$ とする。しかし、上式の定式化において、観測された混合信号の瞬時振幅 $S_k(t)$ と瞬時位相 $\phi_k(t)$ から、四つのパラメータ ($A_k(t), B_k(t), \theta_{1k}(t), \theta_{2k}(t)$) を同時に、かつ一意に求めることはできない。これは二波形分離問題が不良設定の逆問題であることに起因している。そこで、二波形分離モデルでは、Bregmanによって提唱された四つの発見的規則 [2] を表 1.1 に示すような制約条件として用いている。

しかし、これらは定性的なものであるため直接制約条件として利用できない。そこで、これらの制約条件を以下のように定式化して利用している。

制約条件 1 (立上り・立下りの同期) 基本波成分の立上り時刻を T_s 、立下り時刻を T_e とする。このとき、同じ音源出生じた信号成分であれば、高調波成分の立上り $T_{k,on}$ と立下り $T_{k,off}$ は基本波の立上りと立下りに一致しなければならない。すなわち、それぞれの一致の誤差は

$$|T_s - T_{k,on}| \leq \Delta T_s \quad (2.9)$$

$$|T_E - T_{k,\text{off}}| \geq \Delta T_E \quad (2.10)$$

を満たさなければならない。

制約条件 2 (漸近的变化 (多項式近似)) ある区間における瞬時振幅 $A_k(t)$ 、瞬時位相 $\theta_{1k}(t)$ のそれぞれの導関数が、

$$dA_k(t)/dt = C_{k,R}(t) \quad (2.11)$$

$$d\theta_{1k}(t)/dt = D_{k,R}(t) \quad (2.12)$$

で表されるものとする。但し、 $C_{k,R}(t)$ 、 $D_{k,R}(t)$ は、区分的に微分可能な R 次多項式である。このとき $A_k(t)$ 、 $\theta_{1k}(t)$ は、それぞれ、 $A_k(t) = \int C_{k,R}(t)dt + C_{k,0}$ 、 $\theta_{1k}(t) = \int D_{k,R}(t)dt + D_{k,0}$ と表される。

制約条件 3 (漸近的变化 (なめらかさ)) 閉区間 $[t_a, t_b]$ における $A_k(t)$ と $\theta_{1k}(t)$ に対し、定積分

$$\sigma_A = \int_{t_a}^{t_b} [A_k^{(R+1)}(t)]^2 dt \quad (2.13)$$

$$\sigma_\theta = \int_{t_a}^{t_b} [\theta_k^{(R+1)}(t)]^2 dt \quad (2.14)$$

が最小になるとき、 $A_k(t)$ および $\theta_{1k}(t)$ を最もなめらかであるとする。但し、 $A_k(t)$ と $\theta_{1k}(t)$ は、それぞれ、式 (3.) の $C_{k,R}(t)$ と式 (3.) の $D_{k,R}(t)$ で決定された瞬時振幅と瞬時位相である。また、 $A_k^{(R+1)}(t)$ と $\theta_{1k}^{(R+1)}(t)$ はそれぞれ、 $A_k(t)$ と $\theta_{1k}(t)$ の $(R+1)$ 次導関数である。

制約条件 4 (調波関係) 基本周波数を $F_0(t)$ 、高調波の次数を N_{F_0} とする。このとき、調波関係にある信号成分は

$$n \times F_0(t), n = 1, 2, \dots \quad (2.15)$$

の関係を満たさなければならない。

制約条件 5 (振幅包絡 $A_k(t)$ 間の相関) 振幅包絡 $A_k(t)$ は隣接する分析フィルタにおける振幅包絡 $A_l(t)$ に強い相関がなければならない

$$\frac{A_k(t)}{\|A_k(t)\|} \approx \frac{A_l(t)}{\|A_l(t)\|}, \quad k \neq l \quad (2.16)$$

これにより、分離抽出したい信号を拘束することで、 $S_k(t)$ と $\phi_k(t)$ から四つのパラメータを一意に求めることを可能にしている。最後にすべての分析フィルタについて、瞬時振幅 $A_k(t)$ と瞬時位相 $\theta_{1k}(t)$ を求め、式 (2.1) により周波数成分を合成することで目的音 $f_1(t)$ を再合成することができる。ここで、再合成された目的音を $\hat{f}_1(t)$ とする。

以上、本モデルのトップダウン、ボトムアップ処理によって、目的音の選択と重複周波数成分分離の両者が実現する。

第3章 分離抽出モデルの実装

この章では、第二章で提案した選択的分離抽出モデルの各処理ブロックの実装など詳細を述べる。

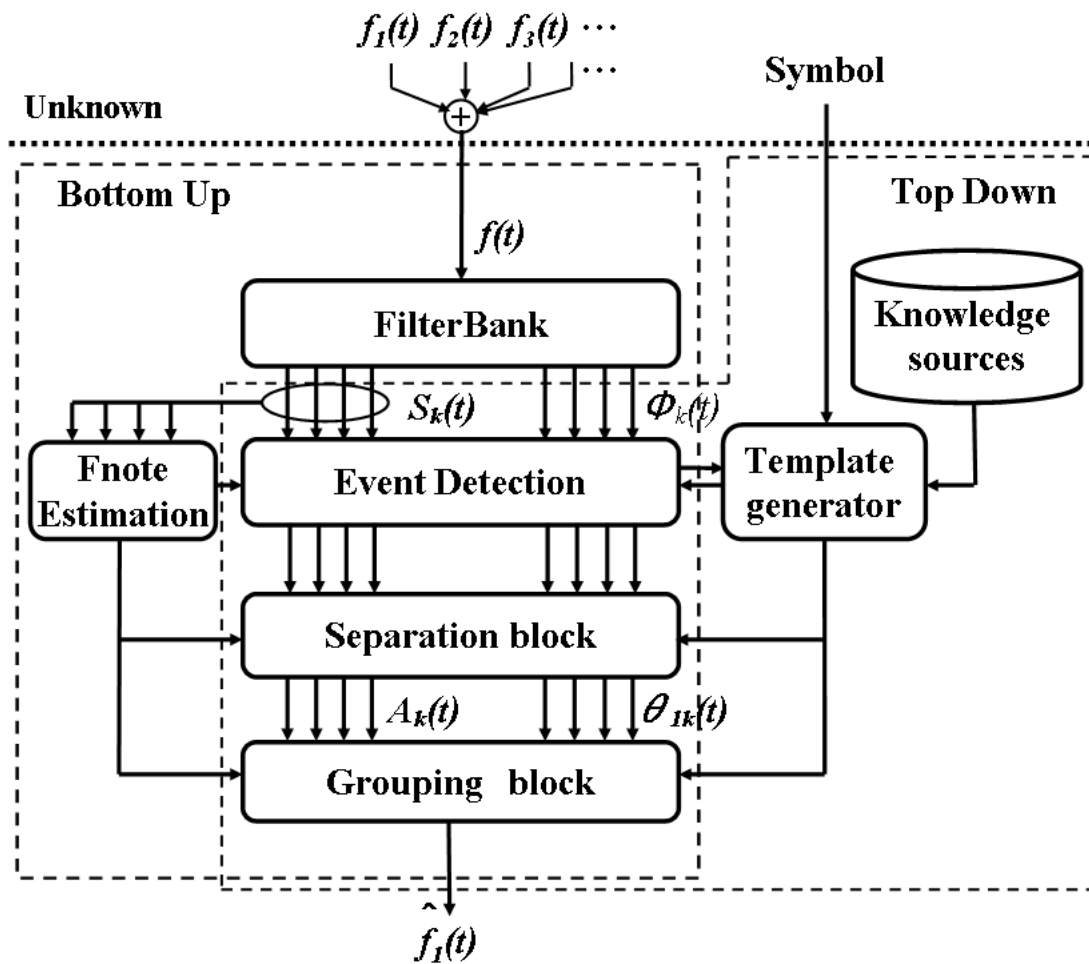


図 3.1: モデルの構成

3.1 周波数分析部

周波数分析部では観測された混合信号 $f(t)$ の瞬時振幅と瞬時位相を抽出し、周波数分析が行われる。本論文で分離の対象とする楽器音は、弦の振動や管の共鳴などにより発生する音であり、調波性が重要な特徴である。そこで、高い周波数領域における調波成分の情報を有効に利用するために、周波数分析部を定帯域幅フィルタバンクとした。また、楽器音はかなり高い周波数成分を含むため、サンプリング周波数を 20 [kHz] とし、解析周波数範囲を 0~10 [kHz] としている。

さらに、今後の処理において混合信号中からそれぞれの楽器音の調波成分などの情報を抽出し利用することを考えた場合、各調波成分の重複を抑えるためにもより狭い周波数帯域幅に分割できるほうが望ましい。そこで、それぞれのフィルタの帯域幅を 20 [Hz] (500 チャンネル) とかなり狭帯域として周波数成分の重なりを抑えている。

表 3.1 にフィルタバンクの仕様を示す。

表 3.1 フィルタバンクの設計仕様

フィルタ数	500
解析周波数範囲	0~10000 Hz
帯域幅	20 Hz
サンプリング周波数	20 kHz

また、このフィルタバンクを利用した楽器音の分析と再合成の一例を図 3.2 に示す。使用した楽器音はフルートの A 音である。この楽器音を分析した結果が図 3.2(b) である。この図は分析によって得られた瞬時振幅 $S_k(t)$ を時間周波数平面で表したもので、その振幅の大きさをグレースケールで表現している。次に、分析した信号を分析フィルタの逆の操作を行う合成フィルタにより再合成した信号が図 3.2(c) である。この再合成信号と原信号の差を雑音とみなす SNR を測定したところ、20.58 dB であった。このことより、本分析合成系の処理では周波数解析のみの処理であればほぼ完全に原信号を復元できることがわかる。

3.2 基本周波数推定部

本論文で対象とする楽器混合音はピッチを知覚するような楽器音 (周期信号) が複数混在する信号である。第 2 章で述べたように、このような信号からの選択的分離抽出を考えた場合、目的音となる可能性をもつ楽器音 (周期信号) が複数存在するため、それらの中から分離対象とする楽器音を選択する必要がある。

目的音の選択的分離抽出することを考えた場合、混合音中に含まれるそれぞれの音源の基本周波数を得ることができれば非常に有効な手がかりとなる。特に、混合音中の複数の音源が時間的に重なりを持つ (同時刻に鳴っている) ような状況からの分離を考えた場合には基本周波数の重要性が明確である。なぜなら、音源同士が時間的に重複している場合

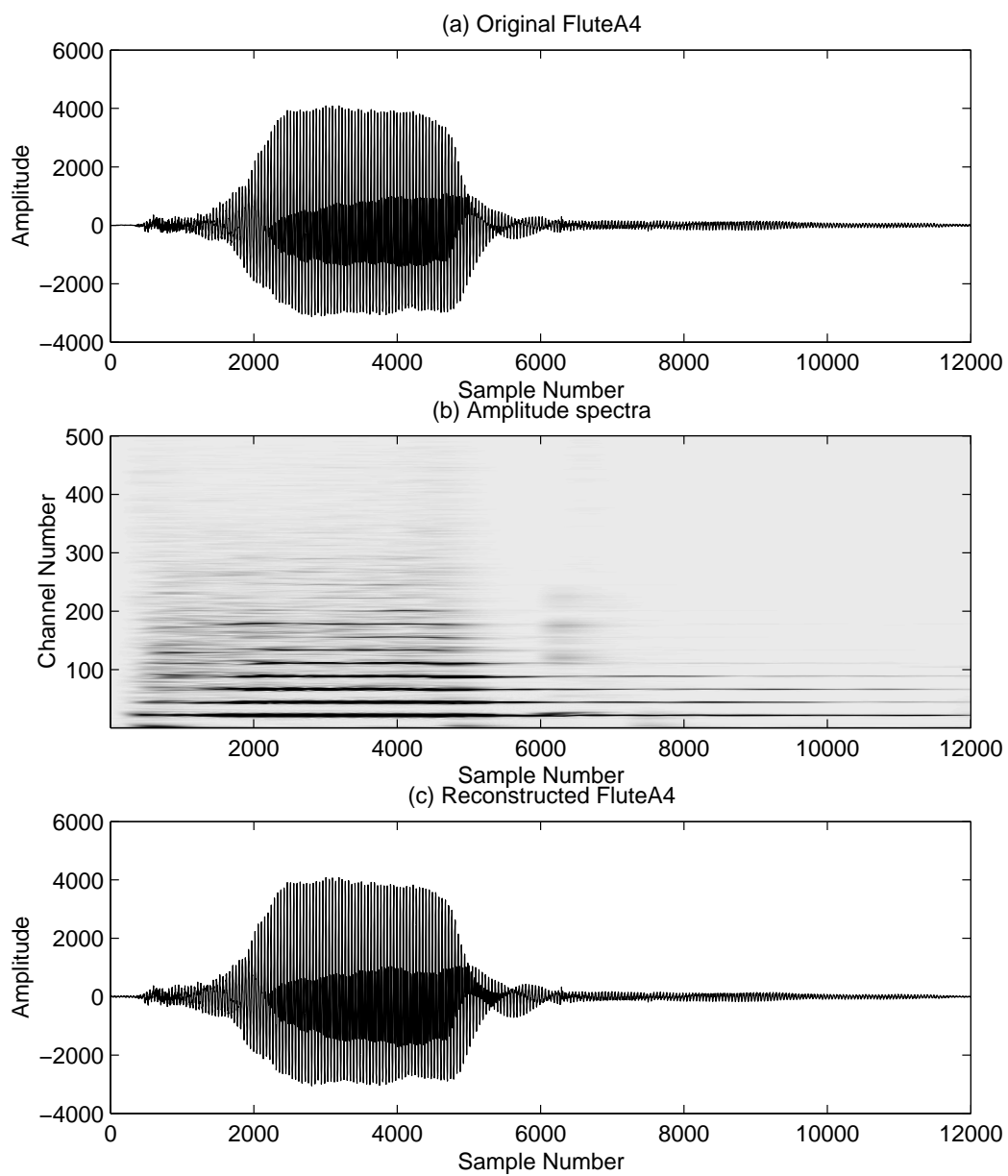


図 3.2: 分析合成系の評価 : (a) 原信号, (b) 瞬時振幅 $S_k(t)$, (c) 再合成信号

には、時間領域の処理のみでそれぞれの音源を分離することは不可能であるのに対して、それぞれの音源の基本周波数を推定できれば信号の調波性などを利用した周波数領域での処理を行うことが可能となるからである。

しかし、音源分離処理の重要な情報である基本周波数を推定する手法としてこれまでに提案されている基本周波数推定法では、単一音を対象とする手法がほとんどである。また、目的音に雑音が付加した状況からの基本周波数推定については数多くの検討がなされているが、音源が複数混合した信号から音源の数とその基本周波数を推定する試みはそれほど多くない。

そこで、提案するモデルの基本周波数推定部では次のような推定法を新たに構築して用いることとした。実際の基本周波数抽出例として、四種類の楽器音フルート (A4)、ピアノ (G3)、ホルン (Eb2)、ヴァイオリン (C3) の混合音からそれぞれの基本周波数を推定する例を示すとともにその処理の流れを述べる。

はじめに、図 3.3 のように混合信号の各時刻における定帯域フィルタバンク出力から周波数軸方向に対する自己相関関数を求める。次に、得られた自己相関関数から最大ピーク一点のみではなく、複数個のピークを検出し基本周波数候補を抽出することで時間的に重なった音源の基本周波数を推定することを目指した。これにより、ある時刻において最もパワーの強い信号の基本周波数のみを推定していた従来の方法に対し、各時刻での信号に含まれる基本周波数候補を複数個得ることができる。各時刻で得られた候補を時間周波数平面上に全て表示したものが図 3.4 である。

次に、図 3.4 のように得られた基本周波数候補の周波数軸に対するヒストグラムを求める。これにより、混合音中の全体にわたって出現する頻度の高い候補がヒストグラムのピークとして得られ、それぞれの音源の基本周波数となる。ただし、この処理で推定される基本周波数には音源の持続時間や時間変化などの情報は含まれていない。このため、3.4 節で述べる目的音選択部では得られた基本周波数と混合音の周波数成分を利用して、それぞれの音源の立上り、立下り時刻を推定し、持続時間を求める処理を行っている。

また、この方法では信号の調波性を利用して基本周波数を推定するため、楽器音の情報を有効に利用できると考えられる。ただし、これには、混合音中の楽器音は通常の演奏法によるものであるとして、楽器音の基本周波数は持続時間中ほぼ一定である、という仮定が含まれている。

3.3 特徴量テンプレート生成部

特徴量テンプレート生成部では、目的音の選択と目的音波形の分離を行う際に利用するテンプレートとして、目的音の調波成分の振幅包絡形状と、各調波成分のパワー比を反映した時間周波数平面上の形状を生成する。特徴量テンプレートの生成には、目的音候補の基本周波数と持続時間、さらに知識源として持っている各楽器の特徴を利用する。

ここで、知識源に蓄えられている各楽器の特徴量とは、それぞれの楽器音の周波数成分を分析することによって得られたものである。図 3.5 にフルート、ピアノ、ヴァイオリン

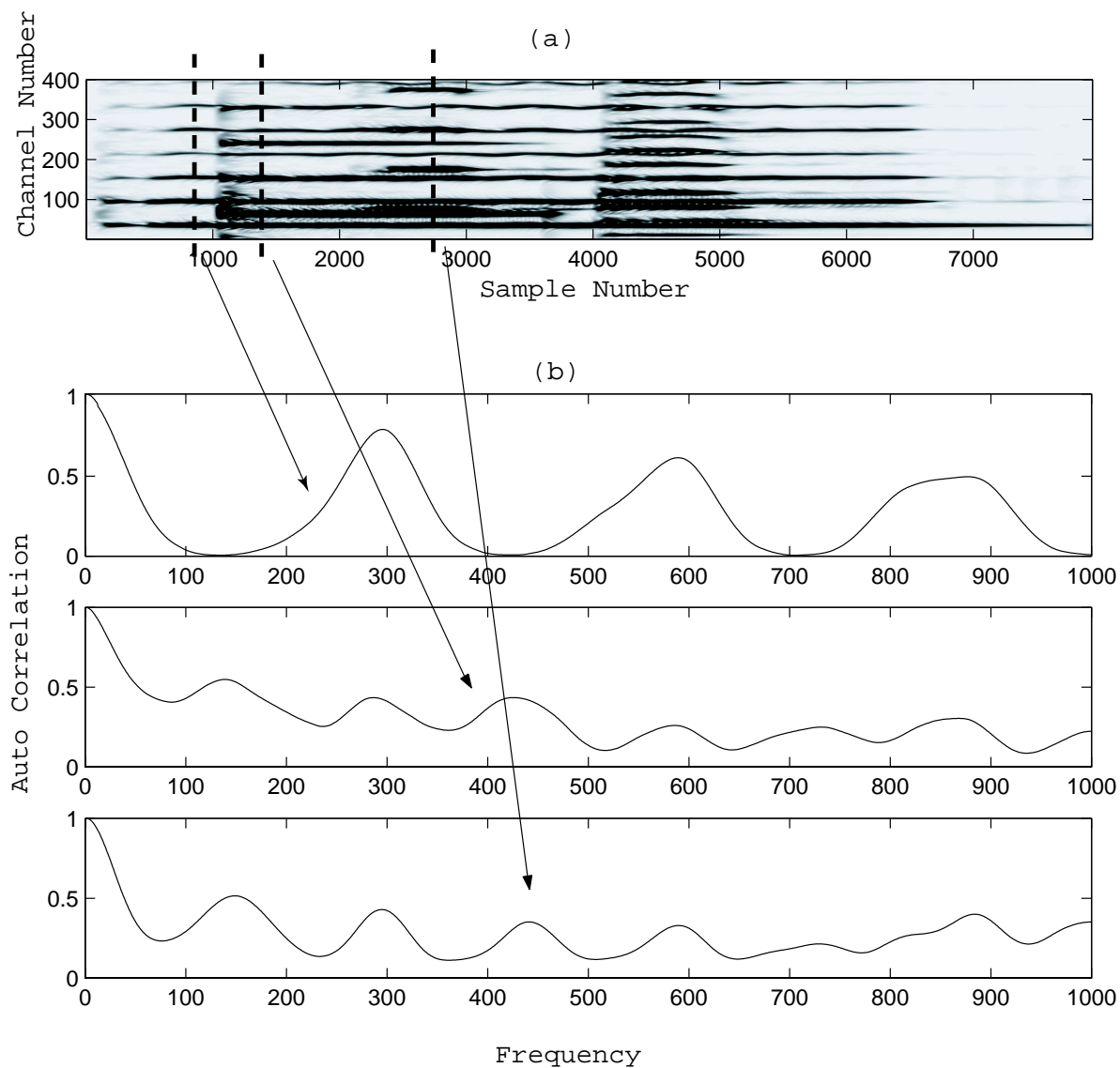


図 3.3: 瞬時振幅 $S_k(t)$ と $S_k(t)$ に対する周波数軸上の自己相関関数 : (a) 瞬時振幅 $S_k(t)$ 、(b) ある時刻における $S_k(t)$ の周波数軸に対する自己相関関数、各時刻で得られた自己相関関数からピークを複数抽出し各ピークでの周波数をそれぞれ基本周波数候補とする

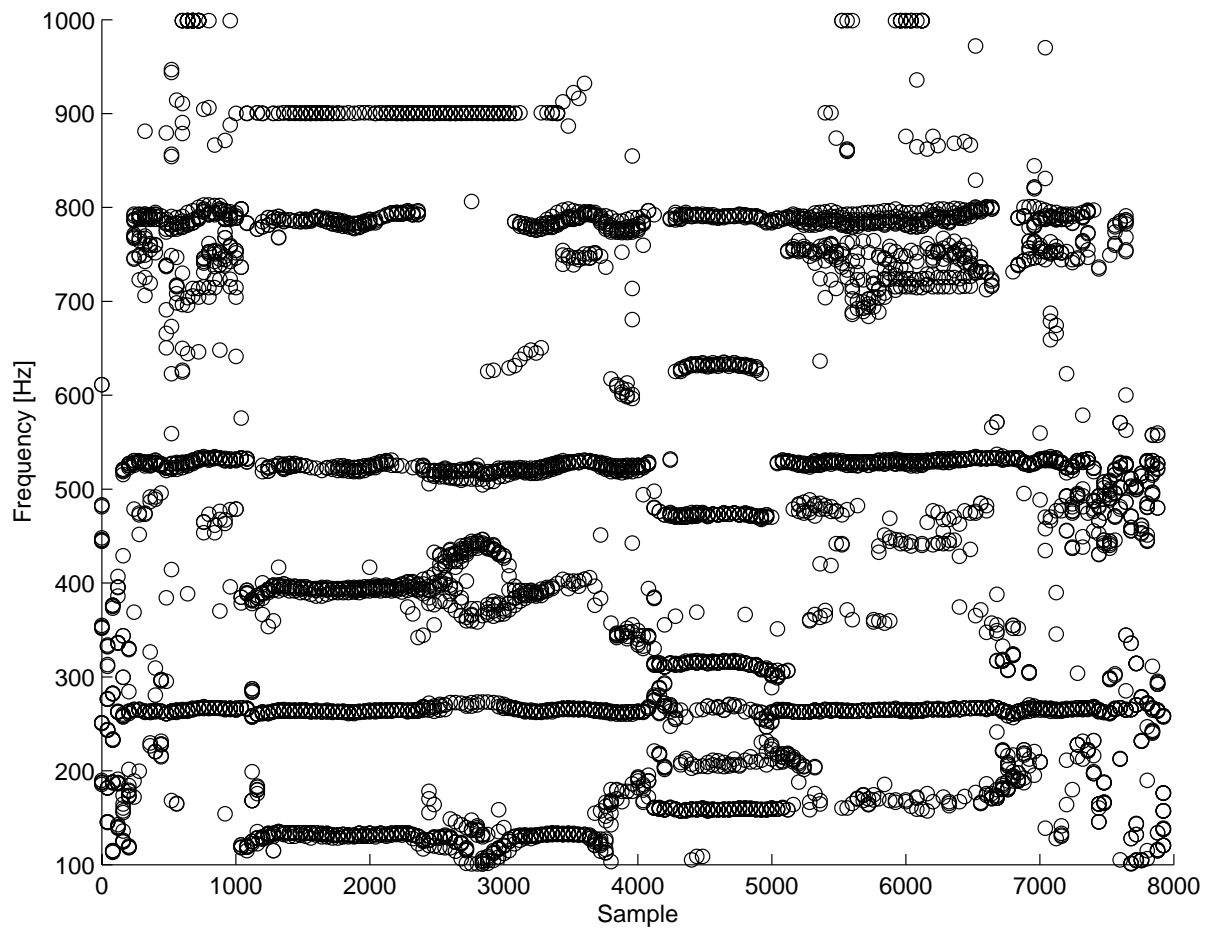


図 3.4: 推定された基本周波数候補：各時刻での周波数方向の自己相関関数のピークを5候補抽出し、時間周波数平面上に表示した図

による単音を本モデルのフィルタバンクにより周波数分析した例を示す。このような時間周波数平面での形状からそれぞれの楽器音を判別するための特徴として以下のようなものが得られる。これらの特徴は楽器音の同定手法 [20, 22] にも利用されているものである。

- フルートなどの木管楽器の特徴
 - 各調波の振幅包絡の定常部分が平坦である。
 - 高調波成分の数が7次以下である（周波数方向での減衰が大きい）。
- ピアノなどの打弦楽器の特徴
 - 各調波の振幅包絡に定常部分がなく、立上り後すぐに減衰する。
 - 高調波成分が比較的多い（10次程度）。
 - 各調波成分の立ち上がりが鋭い。
- ヴァイオリンなどの弦楽器の特徴
 - 各調波の振幅包絡が激しく変動する。
 - 高調波成分の数が多（20次程度）。
 - 高調波成分ではFM成分が見られる。

各楽器のテンプレートは、上記の特徴をもとにまず各調波成分の振幅包絡形状から作成される。このとき、振幅包絡は目的音候補の持続時間に応じて変化させている。その後で、既に得られた基本波の振幅包絡を、楽器によって異なる周波数軸上でのパワー比に従って修正し、各調波成分に並べることでテンプレートを完成させる。

ここではピアノのテンプレート作成例を示す。ピアノ音の振幅包絡に関する特徴は先に示したように、各調波成分の振幅包絡が急激に立上り、その後定常部を持たずにすぐに減衰するという特徴がある。また、楽器音響学の知見から、ピアノなど打弦楽器の振幅包絡における立上り後の減衰は指数関数によって良く近似できることが知られている。そこで、ピアノの各調波成分の振幅包絡を次式によって近似することとした。

$$Tm_{p,k,\text{piano}}(t) = At^{2\alpha-1} \exp\left(-\frac{0.0001}{\alpha}t\right), \quad \alpha = \frac{500-k}{500} \quad (3.1)$$

ここで、 k は本モデルのフィルタバンクチャンネル番号に対応しており、それぞれの k に対応した振幅包絡が得られるようになっている。

上式によってテンプレートの振幅包絡の時間的な変化を表現できたが、つぎに周波数方向のパワー変化を表現することを考える。楽器音の各調波成分のパワー比は各楽器の共鳴の特性に依存すると考えられる。実際に様々な音高のピアノ音を分析した結果から調波成分の周波数に対する変化は次式の指数関数で近似できた。

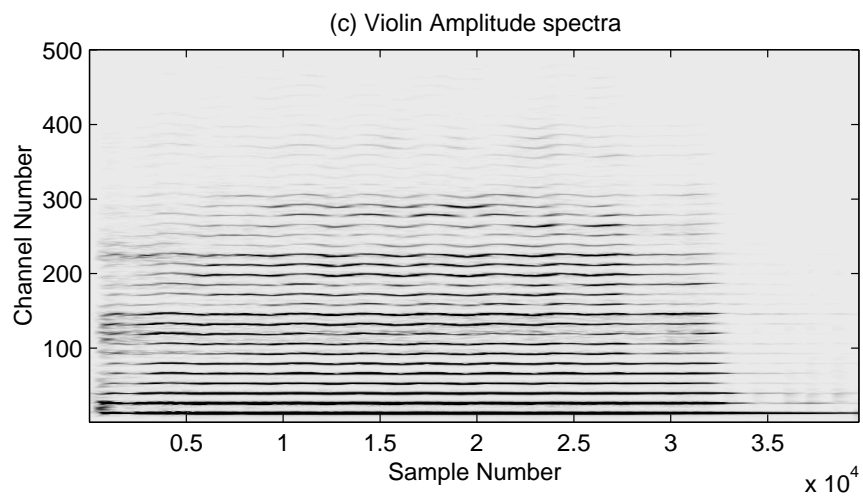
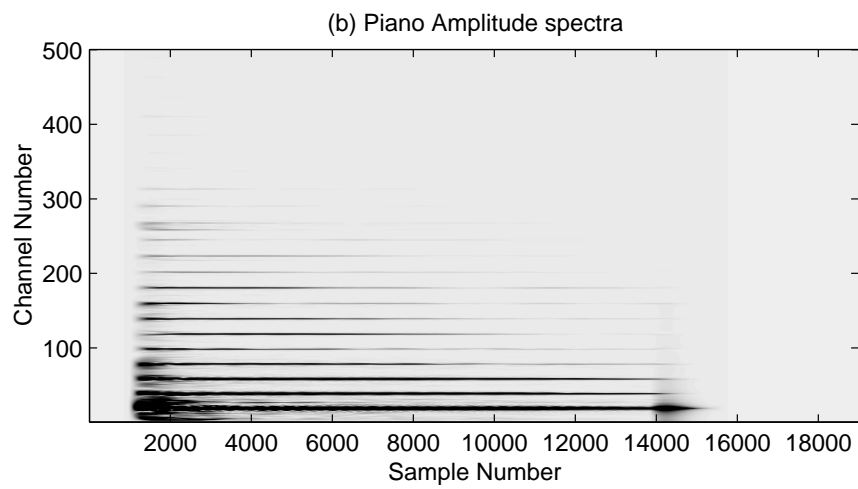
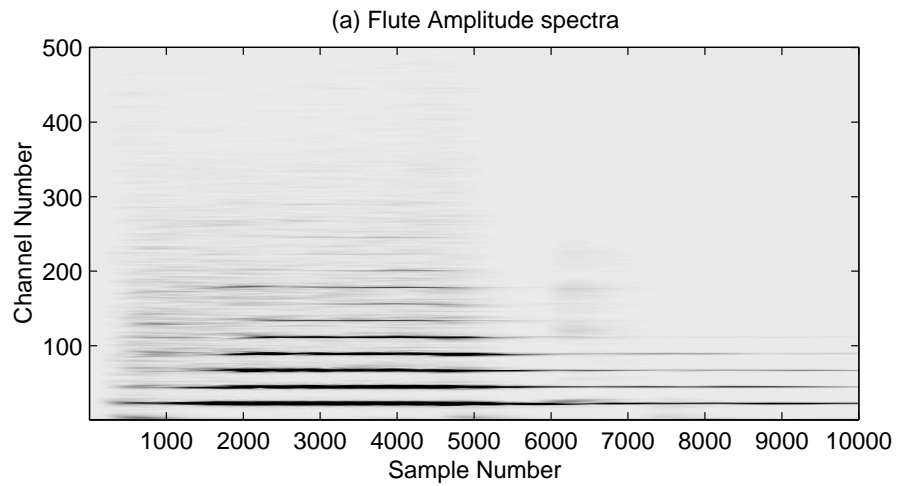


図 3.5: フルート、ピアノ、ヴァイオリン分析例

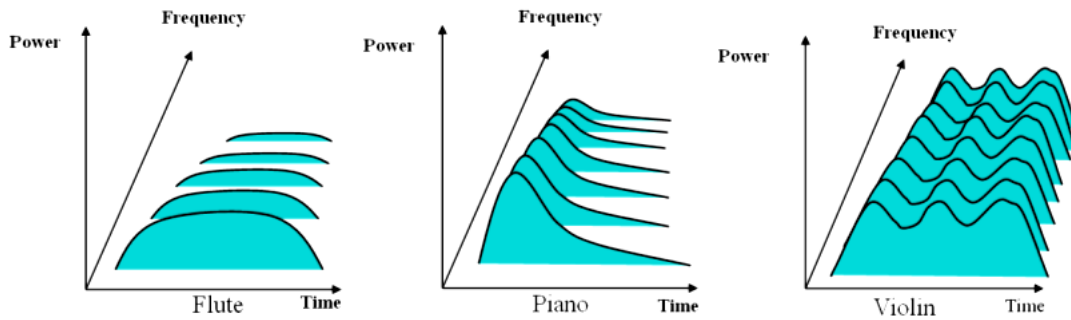


図 3.6: フルート、ピアノ、ヴァイオリンの標準的なテンプレート形状

$$H_{p_{t,piano}}(k) = \exp^{0.04(F_{notek}-k)} \quad (3.2)$$

ここで、 F_{notek} は比較対象となる目的音候補の基本周波数に対応するチャンネル番号である。

よって、先ほど得られた各調波の振幅包絡を (3.2) 式で表現される周波数軸上でのパワー比に従って修正し、時間周波数平面上へ並べることでテンプレートを完成させる。

以上の処理に基づいて作成されたフルート、ピアノ、ヴァイオリンに対する標準的なテンプレートの形状を図 3.6 に示す。

3.4 目的音選択部

目的音選択処理部では、まず、これまでに得た情報から混合音中に含まれるすべての楽器音を目的音候補として抽出し、最終的には各候補と特徴量テンプレートの比較を行うことで候補から目的楽器音成分を選択する。この処理の手順を以下に示し、信号処理の流れを図 3.7 に示す。

1. 推定された基本周波数 (図 3.7:A) の周波数軸に対するヒストグラム (図 3.7:B) から、出現頻度の高い周波数を混合音中の基本周波数候補として抽出する。
2. それぞれの基本周波数候補に対応する分析フィルタの瞬時振幅 $S_k(t)$ から時間領域での立上り、立下りを求める (図 3.7:C)。
3. 瞬時振幅 $S_k(t)$ から 1. で求めた各基本周波数候補の整数倍の成分のみを 2. で求めた持続時間で抽出し目的音候補とする (図 3.7:D,E,F)。
4. 3. で得られた目的音候補と特徴量テンプレート (図 3.7:G) との相関を求め、最も相関の高いものを目的音として出力する (図 3.7:H)。

以上のような基本周波数の整数倍の高調波成分を抽出する処理のみで得られた目的音の周波数成分には、従来のトップダウン処理モデルで問題となっていた他音源の重複周波

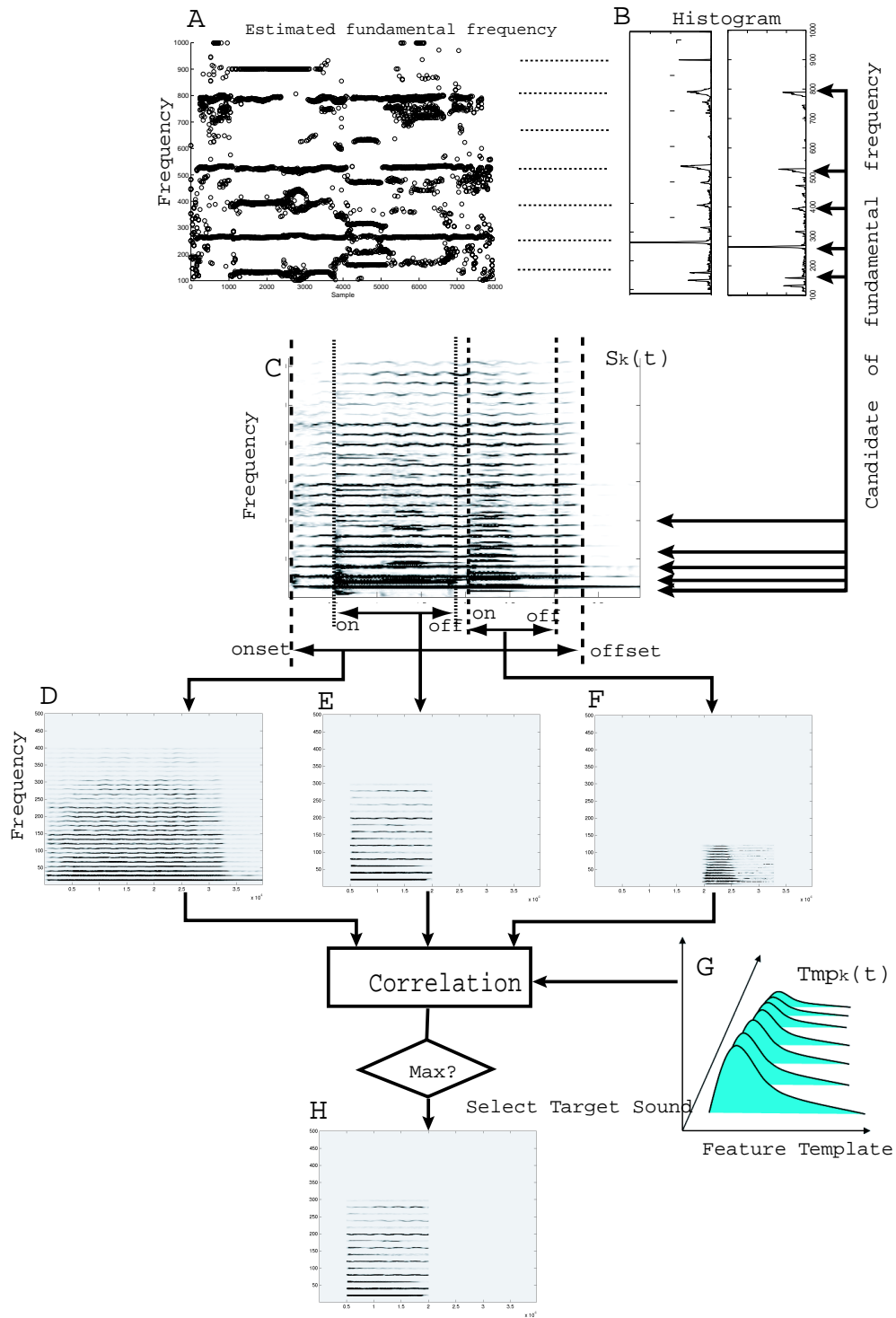


図 3.7: 目的音選択部での信号処理の流れ

表 3.1: 制約条件とその定式化

発見的規則 (Bregman,1993)	制約条件式 (Unoki and Akagi, 1999)
(i) 立上り, 立下りの同期	$ T_S - T_{k,\text{on}} \leq \Delta T_S$ $ T_E - T_{k,\text{off}} \geq \Delta T_E$
(ii) 漸近的变化	
(a) ゆっくりと (多項式近似)	$dA_k(t)/dt = C_{k,R}(t)$ $d\theta_{1k}(t)/dt = D_{k,R}(t)$ $dF_0(t)/dt = E_{0,R}(t)$
(b) なめらかさ	$\sigma_A = \int_{t_a}^{t_b} [A_k^{(R+1)}(t)]^2 dt$ $\sigma_\theta = \int_{t_a}^{t_b} [\theta_k^{(R+1)}(t)]^2 dt$
(iii) 調波関係	$n \times F_0(t), n = 1, 2, \dots$
(iv) 一つの音響事象に生じる多くの変化	$\frac{A_k(t)}{\ A_k(t)\ } \approx \frac{A_l(t)}{\ A_l(t)\ }, k \neq l$

数成分が存在している。そこで、波形分離部において、重複周波数成分の分離を行い、目的音を精度よく抽出する。

3.5 波形分離部

波形分離部では二波形分離モデルを利用して、選択された目的音成分に重複している他音源の周波数成分を分離する処理を行う。

2.3.2 節で述べたように、本モデルのボトムアップ処理に利用する二波形分離モデルでは、Bregman によって提唱された四つの発見的規則を定式化し表 3.1 のような制約条件として用いている。

本モデルの波形分離処理でも二波形分離モデルと同様、これらの発見的規則を制約条件として利用し分離を行うが、その一部に知識源からの情報を適用し分離精度の向上を狙っている。

具体的には、表 3.1 の発見的規則 (iv) の制約条件に注目する。この制約条件は、Bregman の発見的規則 (iv) 一つの音響事象に属する共通の変化に関する規則に対応する制約条件である。二波形分離モデルでは、発見的規則 (iv) を次式のように定式化を行い制約条件としている。

$$\frac{A_k(t)}{\|A_k(t)\|} \approx \frac{A_l(t)}{\|A_l(t)\|}, k \neq l \quad (3.3)$$

この制約条件式は、各調波成分における瞬時振幅 $A_k(t)$ と隣接する調波成分の振幅包絡 $A_l(t)$ の相関が最大となるように、各調波成分における $S_k(t)$ と $\phi_k(t)$ から $A_k(t)$ と $\theta_{1k}(t)$ を求めるものである。

これに対し本モデルでは、この制約条件に知識源からの情報である特徴量テンプレートを利用することを考え、式 (3.3) の $A_l(t)$ を次式のように特徴量テンプレートの振幅包絡 Tmp_k と置き換えたものを制約条件として用いている。

$$\frac{A_k(t)}{Tmp_k(t)} \approx \frac{A_l(t)}{Tmp_l(t)}, \quad k \neq l \quad (3.4)$$

これにより、この制約条件は分離抽出する目的音の瞬時振幅とテンプレートとの相関が最大となるときの $A_k(t)$ と $\theta_{1k}(t)$ を求めるものとなる。このとき、テンプレートとの相関は、式 (3.3) のような大きさを正規化した相関ではなく、大きさも考慮した相関にすることで各調波成分のパワー比による特徴も、分離抽出の処理に反映する形とした。

3.6 グルーピング部

最後にグルーピング部では、波形分離処理によって得られた目的音の瞬時振幅 $A_k(t)$ と瞬時位相 $\theta_{1k}(t)$ をまとめ、それらに対して分析フィルタの逆操作を行うことで、目的音の時間波形を得る。

第4章 分離モデルの評価

本章では、提案する選択的分離抽出モデルを評価するために、様々な混合状態から目的音を分離抽出するシミュレーションを行う。

まず、分離抽出の対象を通常の演奏法による楽器単音として様々な混合の状況から目的の楽器音を選択的に分離抽出し、その分離精度を検証する。

その後で、本モデルがより実環境に近い状況においても有効であることを示すために、数種類の楽器がそれぞれ異なる旋律を同時に演奏している状況から目的楽器による演奏音のみを分離抽出する例を示す。

4.1 シミュレーション条件

4.1.1 評価尺度

モデルによる分離精度を評価するために次のような二つの評価尺度を用意した。一つは、次式に示すような SNR を利用して行った。

$$\text{SNR} = 10 \log_{10} \frac{\int_0^T f_1(t)^2 dt}{\int_0^T (f_1(t) - \hat{f}_1(t))^2 dt} \quad (4.1)$$

この評価尺度は、原信号 $f_1(t)$ と選択的に分離抽出された信号 $\hat{f}_1(t)$ との差を雑音とみなした場合の時間領域での SNR である。この評価尺度を用いる目的は、目的音の分離精度を波形レベルで検証する事である。これにより、二波形分離モデルを利用して分離処理を行う提案法が目的音の瞬時振幅のみでなく瞬時位相の分離も行い、かつ正確に目的音波形を復元する事ができるかを検証することができる。

もう一つは、次式で表される Precision である。

$$\text{Precision} = \frac{1}{T} \int_0^T \left(10 \log_{10} \frac{\sum_{k=1}^K \tilde{A}_k(t)^2 dt}{\sum_{k=1}^T (\tilde{A}_k(t) - A_k(t))^2} \right) dt \quad (4.2)$$

これは、原信号の周波数成分 $\tilde{A}_k(t)$ に対する分離された目的音の周波数成分 $A_k(t)$ の分離精度を時間平均したものである。この評価尺度を用いる目的は、目的音成分に他音源成分が重複した状況においても正確に目的音成分のみを分離できているかを検証することである。

本論文では、上記二つの評価尺度を用いてモデルの分離精度を評価することとした。

4.1.2 比較対象

提案法による分離処理がどの程度有効であるかを明確にするためにその他の手法との比較が必要である。しかし、これまでに提案されている楽音分離モデルは分離結果としてその楽器音の種類と音名による記号表現を出力するものがほとんどである。また、音響信号を出力できるモデルであっても、得られた記号表現を MIDI 音源の演奏情報として利用して仮想的な音響信号を出力するものであるため本モデルとの直接的な比較は困難である。そこで、本モデルによる分離処理の過程で得られる結果を以下のように定義し、二つの処理方法との結果と比較した。

- (a) トップダウン処理のみの結果：目的音の基本周波数の整数倍の成分を単純に抽出し、波形分離処理なしで再合成したもの。
- (b) ボトムアップ処理のみの結果：波形分離部において特徴量テンプレートを利用せずに分離を行った結果。

4.2 楽器単音の分離

ここでは、分離対象とする楽器音を、通常の演奏法による単音として、次のような条件で分離抽出シミュレーションを行った。

- (1) 目的音に白色雑音を付加した信号
- (2) 二つの楽器音を混合した信号
- (3) 四種類の楽器音が混合した信号

まず、(1)のシミュレーションにより、本モデルがこれまでよく考えられてきたような雑音中から目的音を分離抽出する音源分離問題に対しても有効な処理であることを確認する。この条件は、混合信号中に目的音以外の周期信号が存在しないという問題設定であるため、複数の音源から目的音を選択する課題は含まれていない。よって、(1)のシミュレーションでは、雑音が付加された信号からでも目的音の基本周波数を正しく推定可能であるかを確認し、さらに、周波数領域の全帯域に雑音が存在するような状況での分離精度を検証する。

(2)のシミュレーション条件は、選択的分離抽出問題の最も基本的な一例として、二つの楽器音が混合した状況からどちらの音源が目的音であるかを選択し、さらに高調波における重複成分の分離効果を検証するものである。

次に、(3)のシミュレーションにより、四種類の楽器音が混合した状況からの分離抽出を行い、本モデルが選択的分離抽出を可能としているかどうかを検証することで、提案法の有効性を示す。シミュレーションで利用する混合信号は、目的音と雑音の SNR を -10 dB から 20 dB まで 10 dB 刻みで変化させた信号とした。

4.2.1 雑音中の目的音分離抽出

まず、白色雑音が付加された状況からの目的音の分離抽出精度を評価した。この条件では、目的音のすべての調波成分に雑音が存在する状況であるので、波形分離効果を検証できる。また、雑音下からの分離を行う従来の音源分離モデルとの比較も可能である。また、このような白色雑音を付加する状況を実際の音楽音響信号と対応づけて考えた場合、この状況はシンバルなど破裂音を生ずる打楽器を含む信号を模擬するものとして考えられる。

今回は、目的楽器音をフルートとし、3オクターブ中の四種類の音高 (E3,A4,D4,C5) について、それぞれのフルートを雑音中から分離抽出するシミュレーションを行った。

分離抽出の一例を図 4.1 に示す。例えば、図 4.1(a) のフルート (A4 音) に SNR=0 dB とする白色雑音 (図 4.1(b)) が付加されたとき、混合信号は図 4.1(c) となる。観測された混合信号は周波数分析部によりその周波数成分が図 4.1(d) のように得られ、それらから第 3 章で述べたように選択的分離抽出モデルにより目的音フルートの周波数成分が図 4.1(e) のように分離抽出され、最後に再合成することで図 4.1(f) に示すような目的音波形が得られる。この分離例では、分離された目的音波形と原信号の差を雑音とする SNR で 14.85 dB の分離精度が得られており、混合信号の SNR=0 dB であるので本モデルの分離処理によって 14.85 dB の改善があったことになる。

また、四種類の音高について求めた分離精度の平均値を図 4.2 に示す。図 4.2 (上) が SNR を評価尺度に用いた場合の分離精度、図 4.2 (下) が Precision を評価尺度に用いた場合の分離精度である。横軸が目的音に付加された雑音レベルを示す SNR であり、縦軸がそれぞれの雑音レベルからの分離精度を表す。また、本モデルの分離精度と比較するためにトップダウン処理のみの結果と、ボトムアップ処理のみの結果を図中に示す。

これらの結果より、トップダウン処理のみの分離精度と提案モデルでの分離精度を比較することで波形分離処理の効果が確認できる。しかし、入力 SNR が高い場合にはあまり大きな差が見られない。これは、本モデルの周波数分析部で用いているフィルタバンクが、20 [Hz]、500 チャネルとかなり狭帯域で信号を分割していることに起因する。つまり雑音レベルの低い混合信号の場合には、狭帯域フィルタバンクで分割することによって目的音の各調波成分に残る雑音のパワーが非常に少なくなり、トップダウン処理によって調波成分のみを抽出するだけでもかなりの分離精度が得られるためであると考えられる。それに対して、入力 SNR が低い場合には本モデルによる波形分離効果が明確に見られる結果となった。

また、ボトムアップ処理のみの結果と比較すると、すべての雑音レベルで提案法が良い分離精度を得られる結果となっている。このことから、波形分離処理に目的音の情報であるテンプレートを利用することの効果を確認できる。

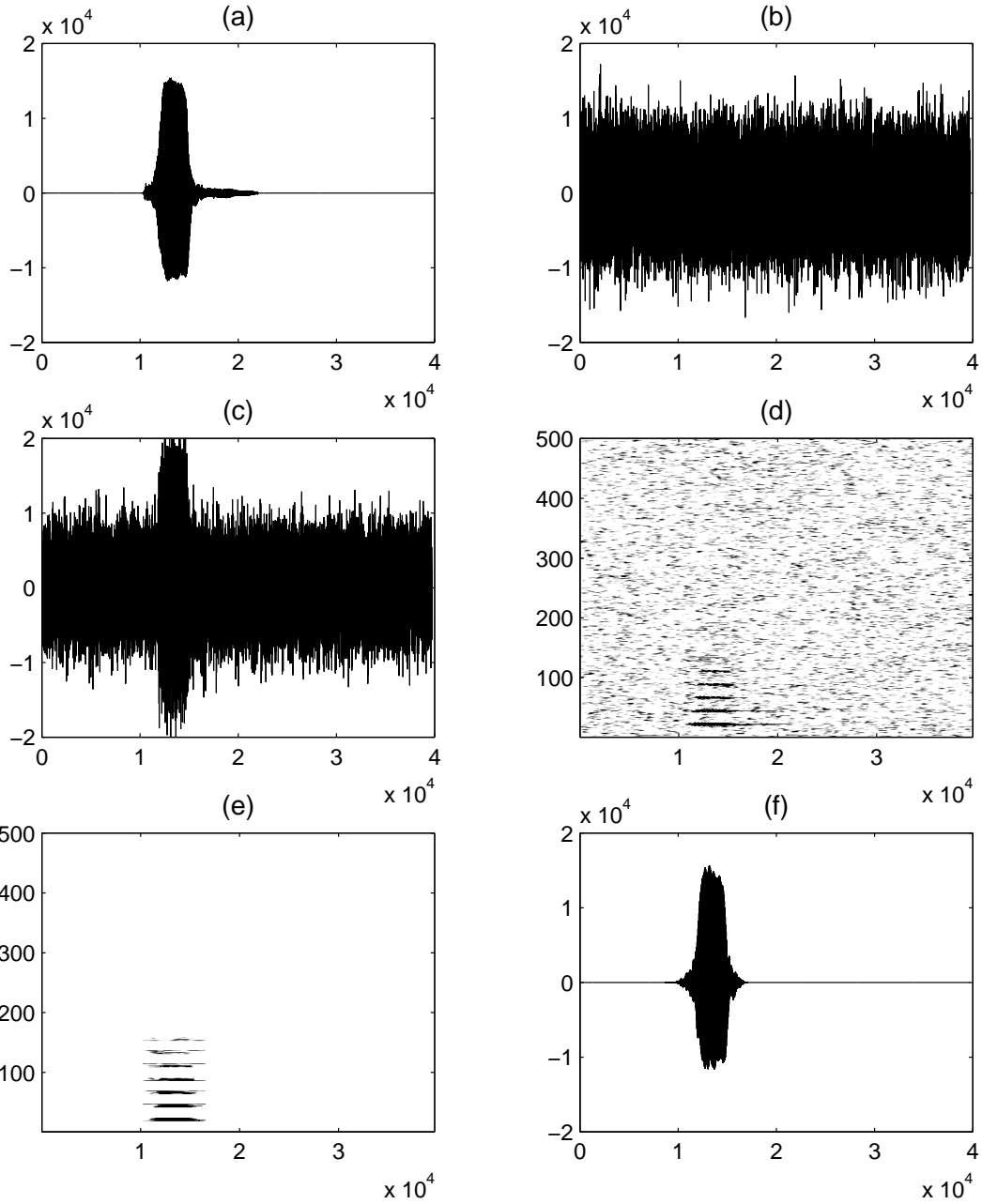


図 4.1: 分離例 (フルーツ (A4)+白色雑音) : (a) 原信号 $f_1(t)$ 、(b) 白色雑音 $f_2(t)$ 、(c) 混合信号 $f(t)$, SNR = 0 dB、(d) 混合信号の周波数成分 $S_k(t)$ 、(e) 分離抽出された目的音フルーツの周波数成分 $A_k(t)$ 、(f) 再合成された目的音フルーツの波形 $\hat{f}_1(t)$

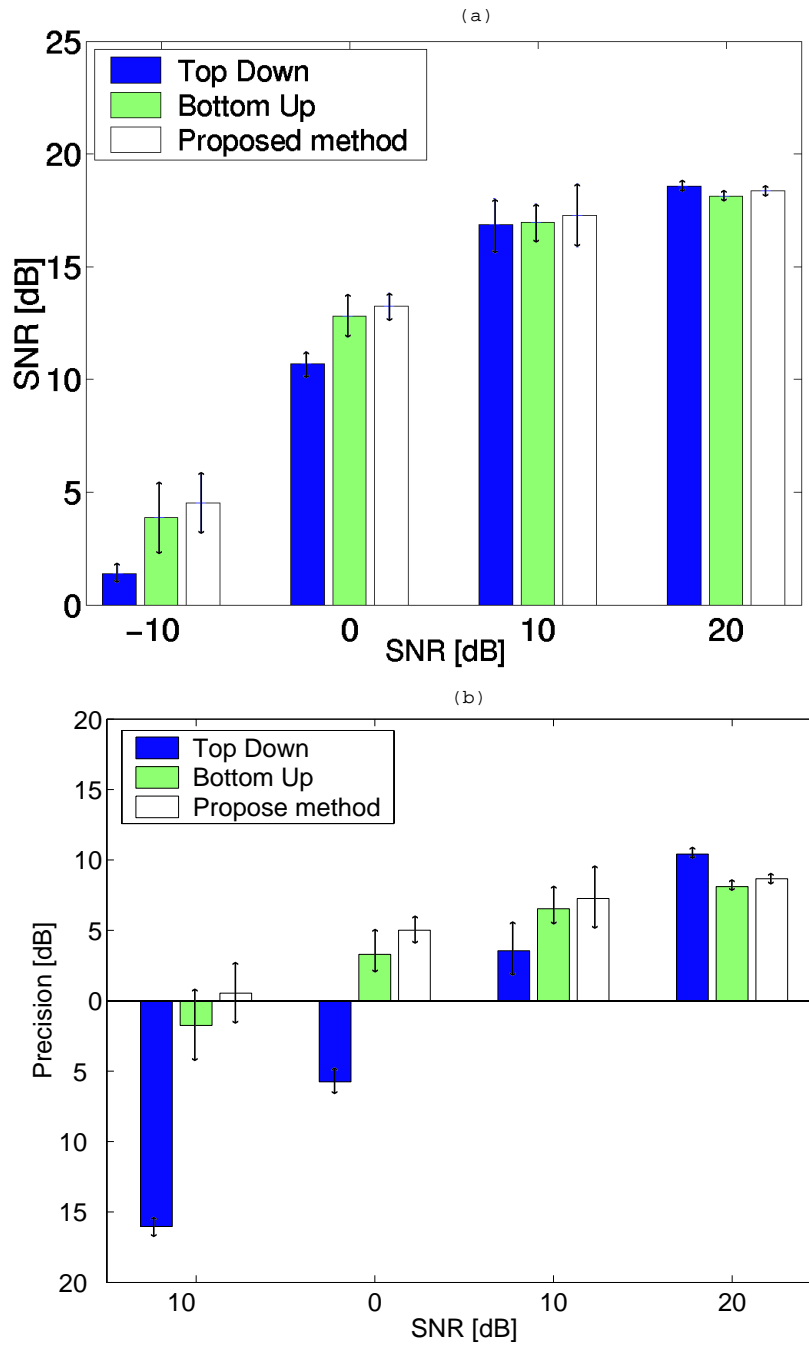


図 4.2: 雑音中からの目的音分離抽出結果 : (a)SNR (b)Precision

4.2.2 二楽器混合音からの分離抽出

次に二種類の楽器音が混合された信号を用いて分離抽出シミュレーションを行った。この混合信号では信号中に目的音候補となる音源が二つだけ存在する状況であり、最も基本的な選択的な分離抽出問題である。この混合信号を用いたシミュレーションにより、本モデルが二つの目的音候補からどちらが目的楽器音であるかを選択し、その成分のみを分離抽出することができるかを検証する。

また、これまでに提案されている楽器音の分離モデルでは、それぞれの楽器音の周波数成分が倍音関係となる場合に生ずる重複周波数成分の分離が行われていないという問題があり、その後の楽器音同定処理などで認識率低下の原因となっていた。それに対して本モデルでは、波形分離部において重複周波数成分の分離を行っている。このシミュレーションでは、本モデルの波形分離部において楽器音同士の重複周波数成分を分離可能であるかを検証する。

混合する楽器音はヴァイオリン (C4) とフルート (A4) とし、目的音をフルートとした。分離抽出の一例を図 4.3 に示す。ここで、図 4.3(a) に示す目的音フルート $f_1(t)$ に対して図 4.3(b) のヴァイオリンを雑音 $f_2(t)$ として SNR=0 dB で混合した信号 $f(t)$ が図 4.3(c) となる。観測された混合信号は周波数分析部によりその周波数成分 $S_k(t)$ が図 4.1(d) のように得られ、それらから第 3 章で述べた選択的分離抽出モデルにより目的音フルートの周波数成分が図 4.1(e) のように分離抽出され、最後に再合成することで図 4.1(f) に示すような目的音波形が得られる。

この分離例では、分離された目的音波形と原信号の差を雑音とする SNR で 16.28 dB の分離精度が得られており、混合信号の SNR=0 dB であるので本モデルの分離処理によって 16.28 dB の改善があったことになる。その他の雑音レベルでの分離抽出結果を図 4.4 に示す。図 4.4 (上) が SNR を評価尺度に用いた場合の分離精度、図 4.4 (下) が Precision を評価尺度に用いた場合の分離精度である。

この結果、混合音中に二つの目的音候補が存在するという基本的な選択的分離抽出問題において目的音選択処理部での処理により目的音を正しく選択できることが確認できた。

また、この二つ音階の組み合わせでは目的音であるフルートの高調波成分 3,6, …次にヴァイオリンの高調波成分が重複する状況である。このような場合、重複周波数成分の分離を行わない従来のモデルでは、図 4.4 のトップダウン処理のみの結果のように入力 SNR が悪くなるに従って分離精度が低下していた。それに対して本モデルの分離結果は、波形分離部での処理による分離効果から SNR、Precision とともに分離精度が改善されている。雑音中からの分離結果と同様、特に入力 SNR が低い場合には大きな効果が見られる。例えば、入力 SNR=10 dB の場合にはトップダウン処理のみの結果と比較して SNR では 4.03 dB、Precision では 14.6 dB の改善効果があった。

このことより本モデルの波形分離処理は、白色雑音などの雑音との重複周波数成分のみでなく、楽器音同士の重複周波数成分の分離に対しても有効な処理であることがわかった。

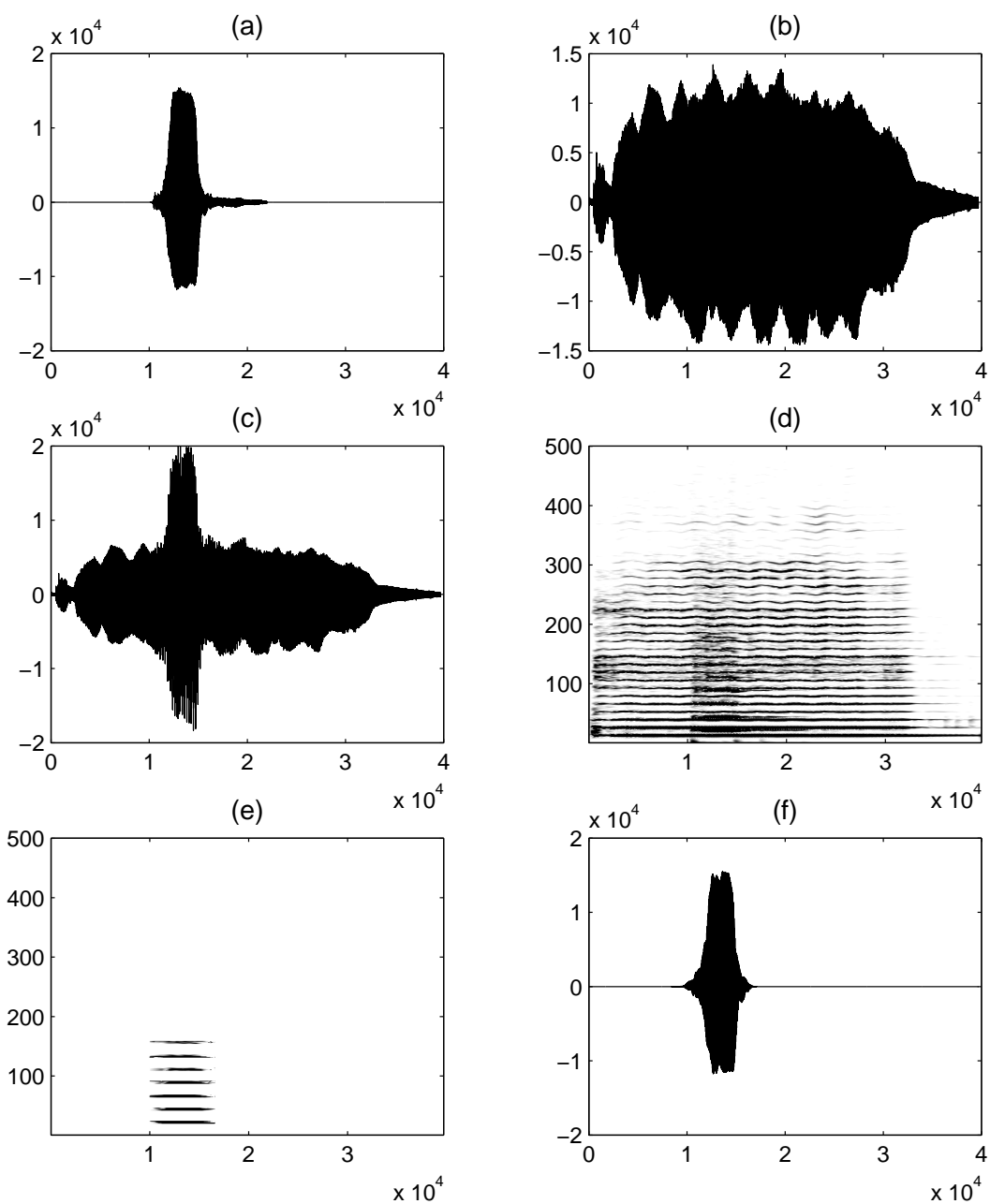


図 4.3: 分離例 (フルート (A4)+ヴァイオリン (C3)) : (a) フルード $f_1(t)$ 、(b) ヴァイオリン (C3) $f_2(t)$ 、(c) 混合信号 $f(t)$, SNR = 0 dB、(d) 混合信号の周波数成分 $S_k(t)$ 、(e) 分離抽出された目的音フルートの周波数成分 $A_k(t)$ 、(f) 再合成された目的音フルート (A4) の波形 $\hat{f}_1(t)$

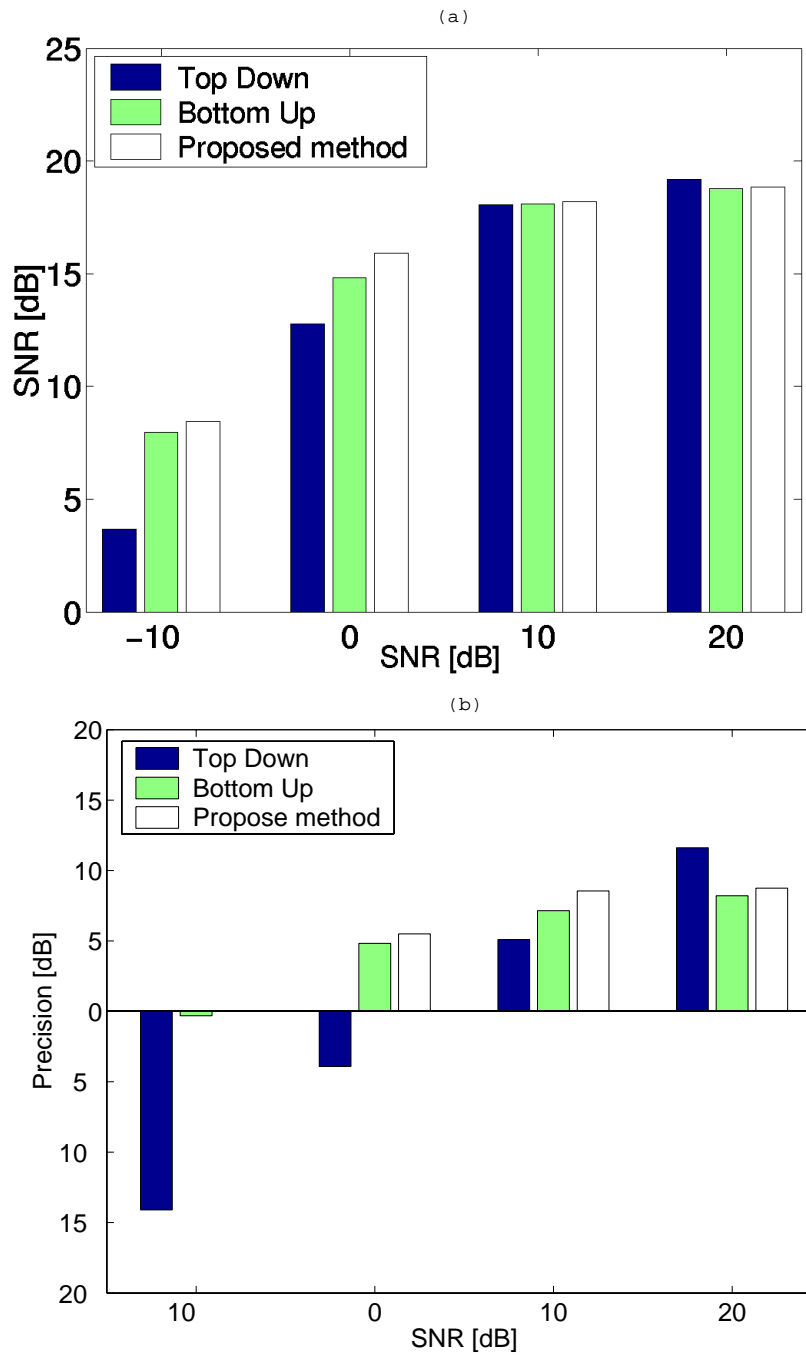


図 4.4: 二楽器混合音からの目的音の分離抽出結果 : (a)SNR (b)Precision

4.2.3 複数楽器混合音からの選択的分離抽出

次に、四種類の楽器音が混合した信号から、特定の楽器音を選択的に分離抽出するシミュレーションを行った。この状況は、四種類の楽器音が混在する音から、聴きたい楽器の音に注目してその音を聴き分けるという行為を模擬するものである。

混合に用いた楽器音はフルート (A4)、ヴァイオリン (C4)、ピアノ (B3)、ホルン (E ♭ 2) の四種類の楽器音である。

分離処理の一例としてフルートを目的音として選択的分離抽出を行った際の信号処理の流れを図 4.5 に示す。混合信号は、目的音フルートに対してその他の楽器音を雑音とみなし SNR=0 dB となるように混合したものである。

この分離例では、目的音フルートを SNR=8.38 dB の分離精度で分離することができた。混合信号は SNR=0 dB であるので、本モデルの分離処理によって 8.38 dB の改善があったことになる。その他の混合状況での分離抽出結果を図 4.6 に示す。図 4.6(a) が SNR による分離精度であり、図 4.6(b) が Precision による分離精度を表している。また、フルート以外の三つの楽器音を目的音として同様に選択的分離抽出を行った結果を図 4.7、図 4.8、図 4.9 にそれぞれ示す。

これらの結果から、本モデルの分離抽出結果とトップダウン処理のみの結果と比較して、SNR、Precision いずれの評価尺度においても本モデルでの分離処理による分離精度改善が見られ、有効性が確認できる。しかし、これまでの結果と同様に雑音レベルが低い状況からの分離抽出精度に注目した場合に、SNR でトップダウン処理のみの結果と大差がない、もしくは若干劣る結果となっている。だが、Precision の値に注目すると、このような状況でもトップダウン処理の値よりも良い値を得ている場合が多く見られる。このことより、波形分離処理において振幅包絡 $A_k(t)$ については正確に分離できているにもかかわらず、位相成分 $\theta_{1k}(t)$ の分離精度が悪いことが原因で、SNR による波形レベルの評価では良い結果が得られていないと考えられる。

しかし、これまでのシミュレーション結果と同様に SNR=-10 dB のような悪条件からの分離抽出では、トップダウン処理のみの結果と比較して大きな改善効果が見られる。

このシミュレーションにより、本モデルによって複数の楽器音が混合した信号中から目的楽器音を正しく選択し、高い精度で分離できることが確認された。

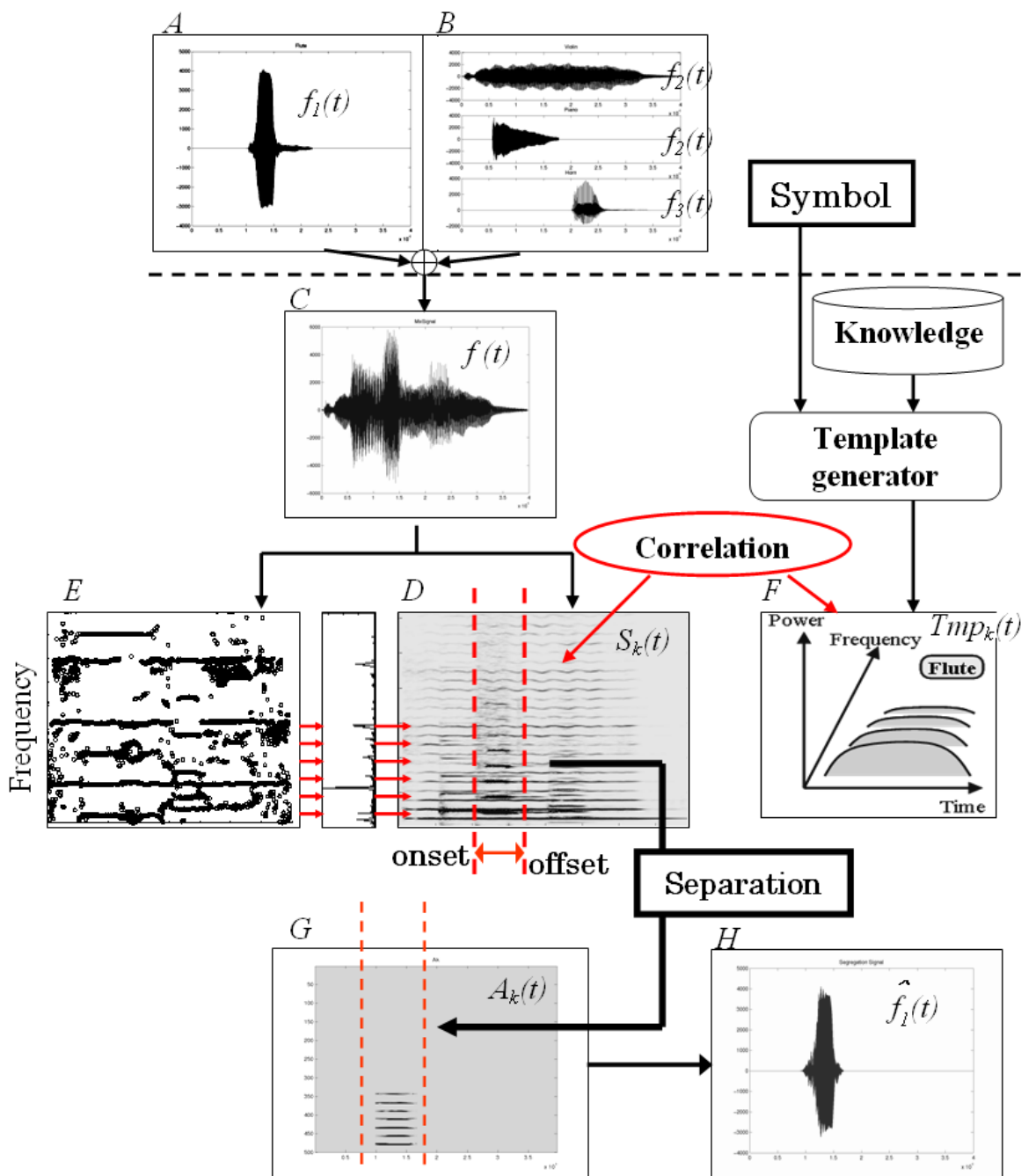


図 4.5: 四種類の楽器音からの目的音分離例(フルート) (A:目的楽器音 $f_1(t)$ 、B:その他の楽器音 $f_2(t), f_3(t), f_4(t)$ 、C:混合音 $f(t)$ 、D:瞬時振幅 $S_k(t)$ 、E:基本周波数 $Fnote$ 、F:特徴量テンプレート $Tmp_k(t)$ 、G:分離された目的音成分 $A_k(t)$ 、H:分離された目的音 $\hat{f}_1(t)$)

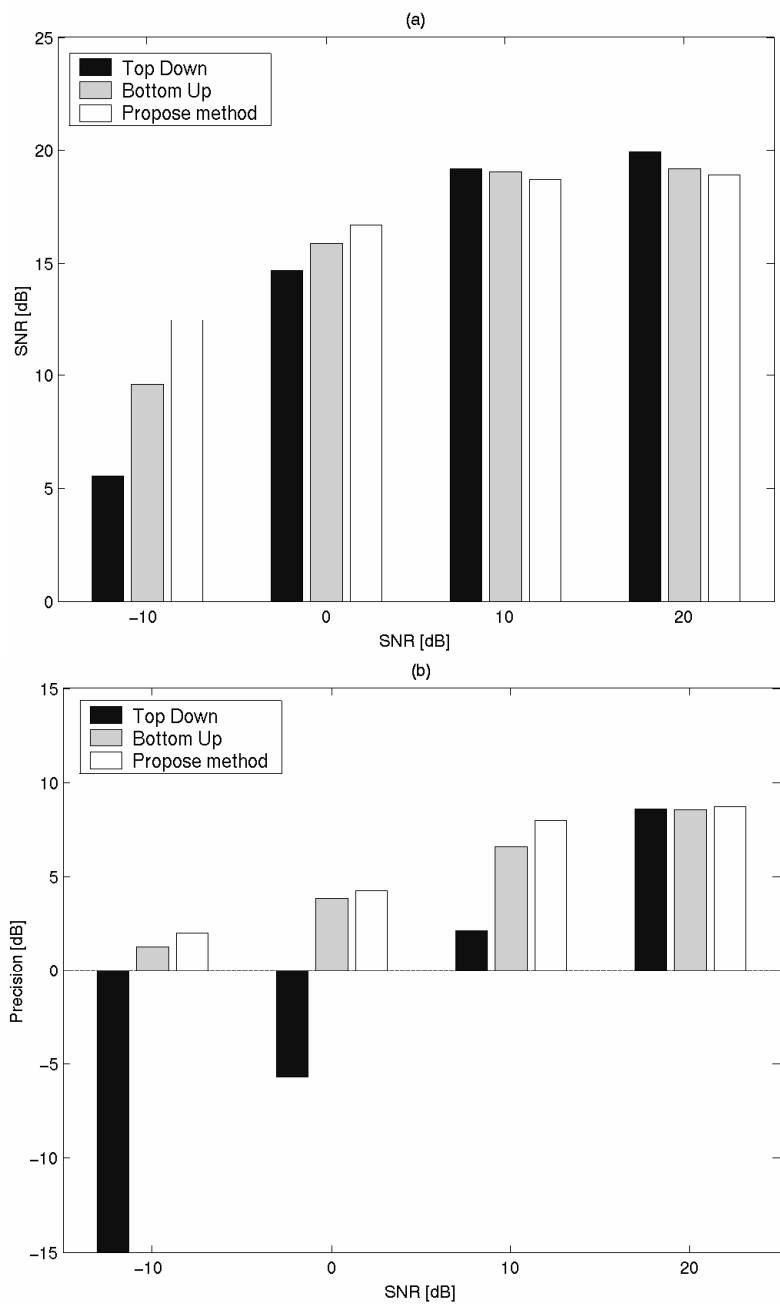


図 4.6: 四種類の楽器混合音からの分離精度 (フルート) : (a)SNR (a)Precision

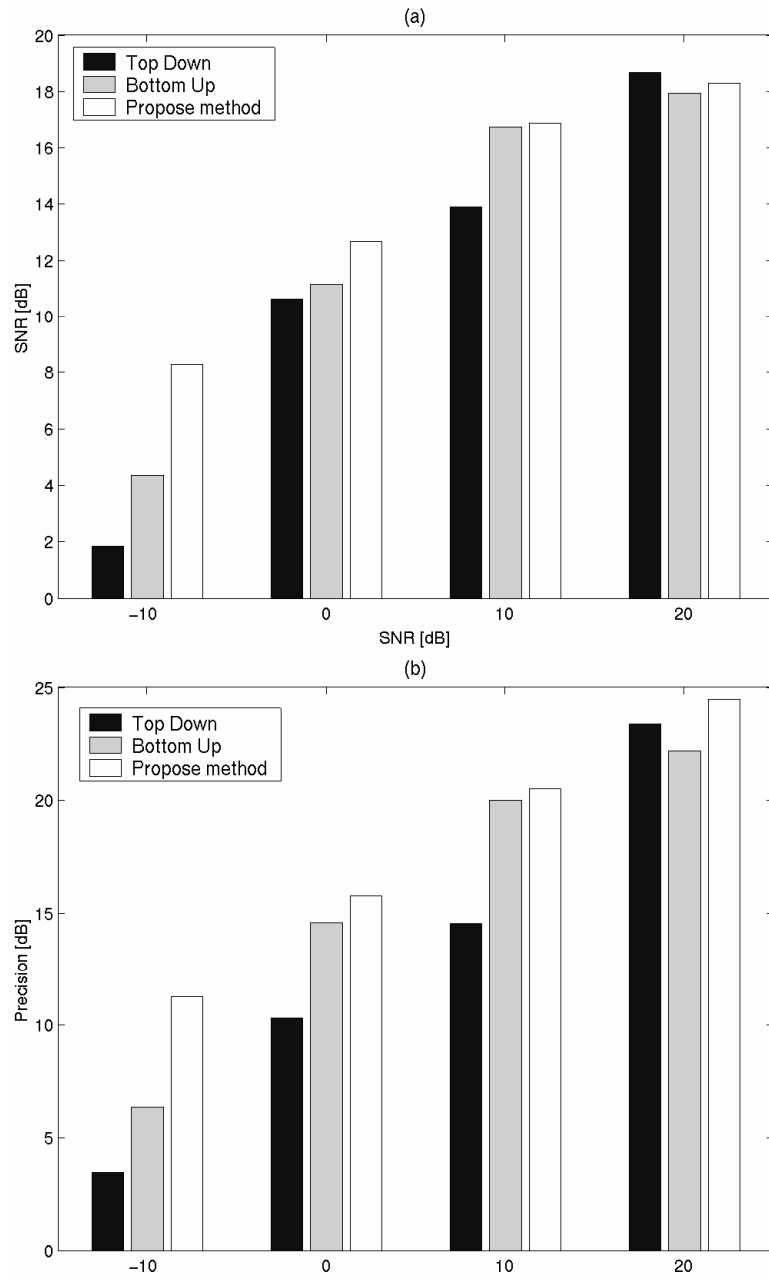


図 4.7: 四種類の楽器混合音からの分離精度 (ピアノ) : (a)SNR (b)Precision

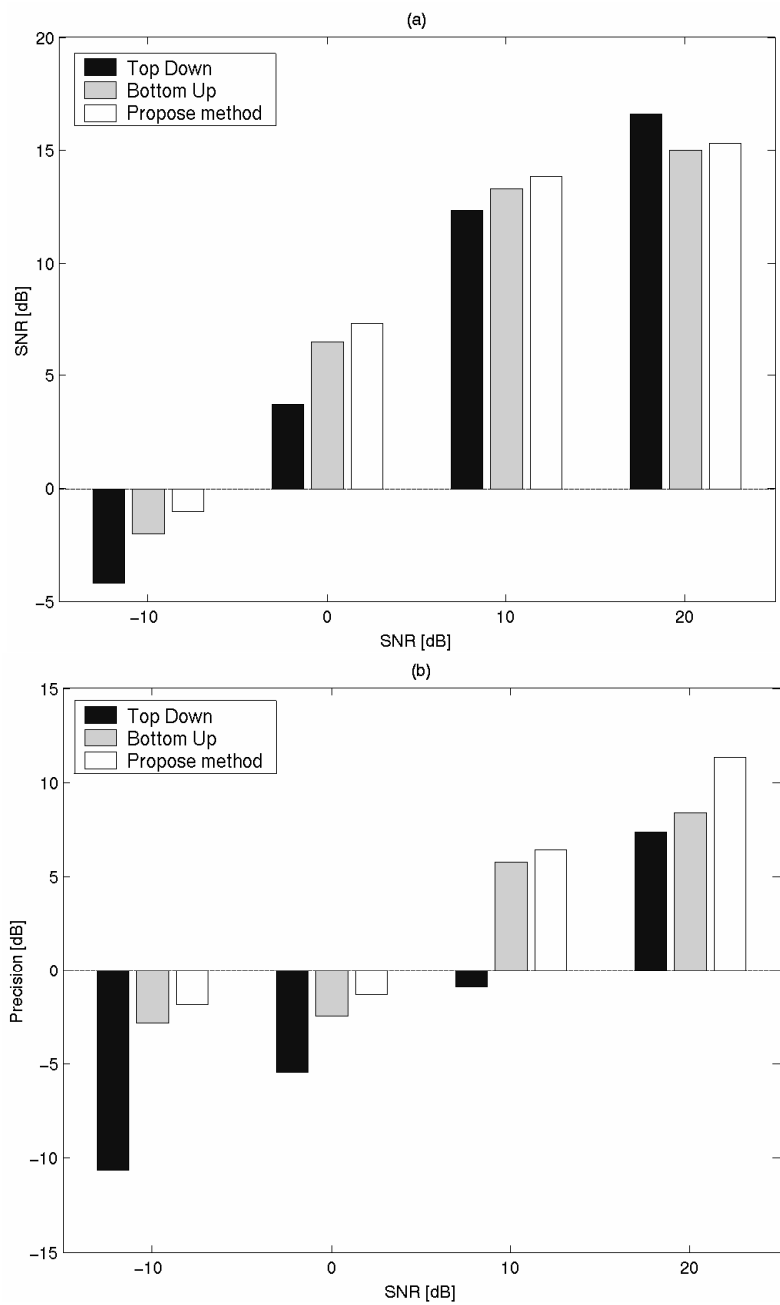


図 4.8: 四種類の楽器混合音からの分離精度 (ホルン) : (a)SNR (b)Precision

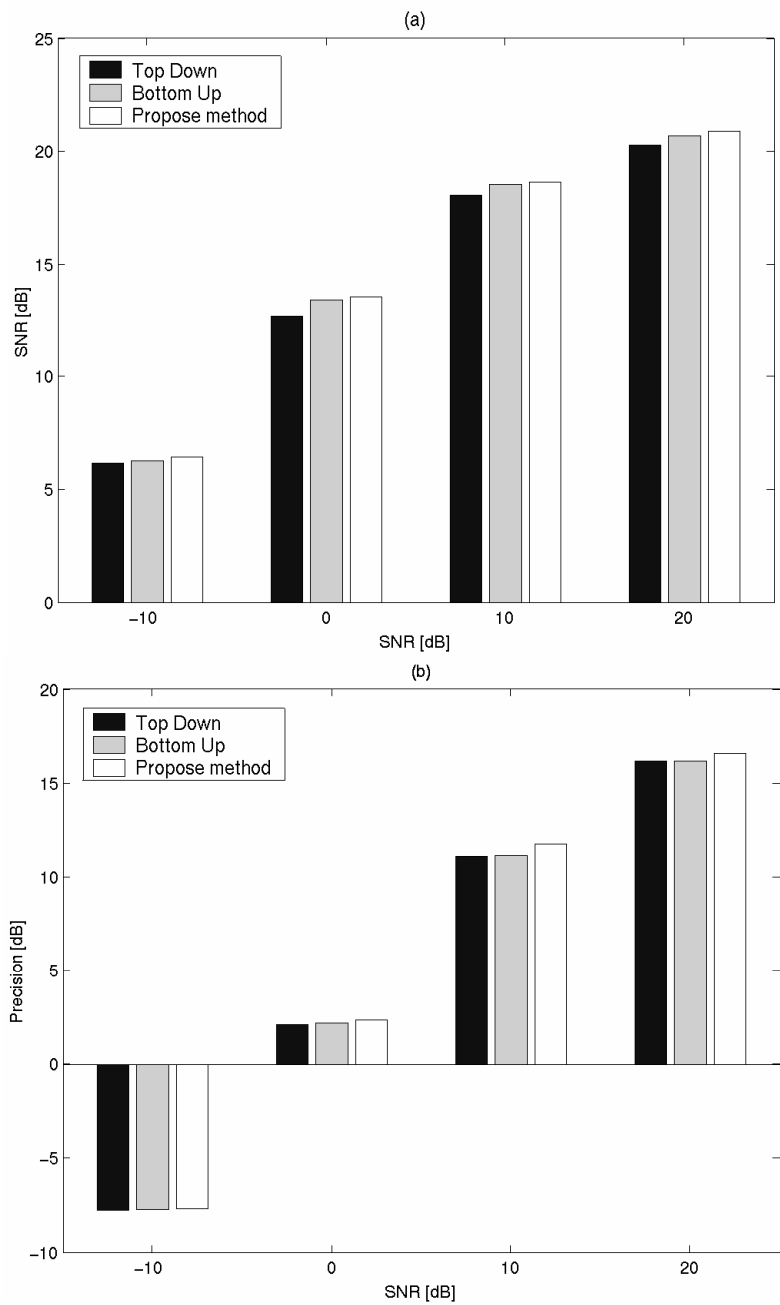


図 4.9: 四種類の楽器混合音からの分離精度 (ヴァイオリン) : (a)SNR (b)Precision

4.3 楽器演奏音の分離抽出

最後に、提案する選択的分離抽出モデルが楽器の単音のみでなく、より実環境に近い状況へも応用可能であることを示すために、旋律を演奏する楽器音の分離抽出を行った。

ここでは、三種類の楽器がそれぞれ異なる旋律を同時に演奏する状況から、ある楽器の旋律に注目し、その楽器による演奏のみを選択的に分離抽出する。我々がこのような複数楽器音による演奏から目的楽器音の演奏を聴こうとする状況を考えると、演奏中の主旋律を担当する楽器音などは比較的容易に聴き分けられるであろう。また、音楽演奏以外では、市街地などの路上において様々な音が渦巻く環境においても自分の携帯の着信メロディなどは容易に聴き分けられるといった例があげられる。

このようなことから、旋律をとともなう音の聴き分けにはその音源の音響的特徴のみでなく、旋律のつながり(メロディ)といった情報を有効に利用していると考えられる。そこで、このシミュレーションではモデルに与える知識として楽器音の音響的特徴以外に、メロディに関する知識を新たに仮定し、追加することとした。具体的には、楽譜に相当するような情報を知識に持つと仮定し、注目する楽器演奏を構成する単音の音高と、その発音順序を与える。従って、モデルとしては、混合信号中の目的音の基本周波数と大まかな発音時刻が利用できることになる。ただし、旋律を構成する楽器単音が実際に存在する時刻やその持続時間は未知であるため、目的音選択部で推定が必要である。

以上のような仮定のもとで次のような混合の条件から目的楽器の演奏音を分離抽出した。混合の条件を表 4.1 に示す。

表 4.1: 楽器演奏音の混合に用いた楽器とその旋律

	楽器	演奏曲名	単音数 (音名)
$f_1(t)$	ピアノ	チューリップ	6音 (CDECDE)
$f_2(t)$	フルート	きらきら星	7音 (CCGGAAG)
$f_3(t)$	ヴァイオリン	チョウチョ	6音 (GEEFEE)
$f_4(t)$	白色雑音	-	-

ここで、三種類の楽器演奏音に加えて、実環境での雑音を表現するために白色雑音を加えることとした。また、このような白色雑音は音楽音響信号においてもシンバルなどの打楽器音で発生する音を模擬すると考えることができる。

分離抽出の一例を図 4.10 に示す。この例では図 4.10(a) のようなピアノによる演奏音 $f_1(t)$ を目的音とし、これに $\text{SNR} = 0 \text{ dB}$ となるようにその他の楽器演奏音 $f_2(t)$, $f_3(t)$ と白色雑音 $f_4(t)$ を付加し、図 4.10(b) に示すような混合音 $f(t)$ からの分離抽出を行った。

モデルによる分離処理は旋律を構成する楽器単音ごとに行われる。まず、混合音中の初めにあらわれるピアノ単音の位置を推定し、その分離を行う。その後、順に楽譜情報と混合信号からの情報から次の目的楽器による演奏音の時間的位置を特定し分離を行う。最終

的に得られるそれぞれの楽器単音を足しあわせ、連続した演奏音を図 4.10(f) のように得ることができる。

分離された 6 つの単音それぞれについての分離精度の平均を求めた結果、SNR で約 12 dB の改善が得られ、Precision では約 16 dB 程度の改善が得られた。また、トップダウン処理のみの結果と比較して、SNR で約 1.5 dB、Precision で約 2 dB 程度の優位性が確認できた。

これらの結果より、本モデルは旋律を持った楽器演奏音の分離に対しても応用できる可能性があることを示すことができた。

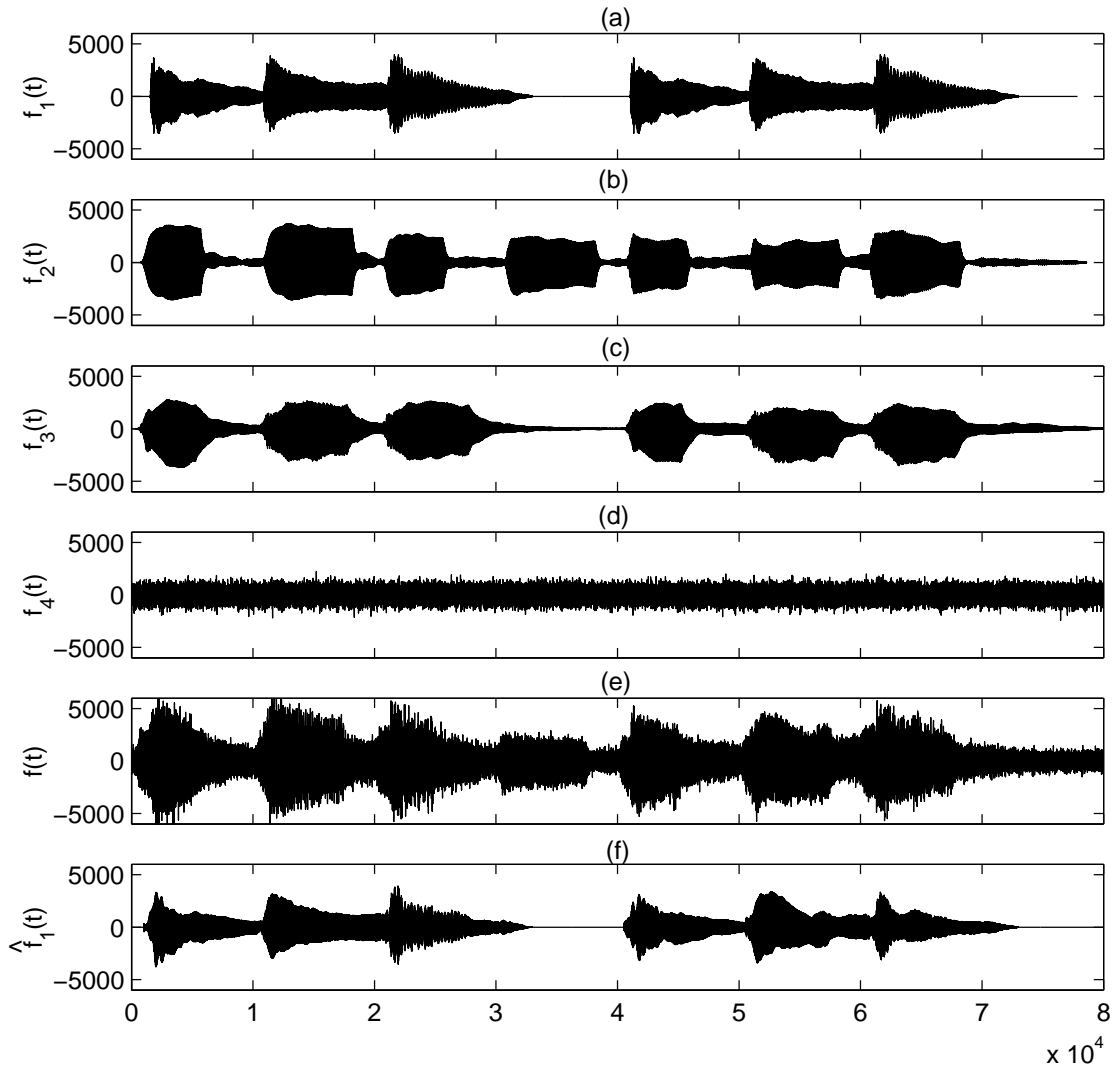


図 4.10: 楽器演奏音の分離抽出例 : (a) ピアノ $f_1(t)$ (目的演奏音), (b) フルート $f_2(t)$, (c) ヴァイオリン $f_3(t)$, (d) 白色雑音 $f_4(t)$, (e) 混合音 $f(t)$ (ピアノ+フルート+ヴァイオリン+白色雑音), (f) 分離された目的演奏音 $\hat{f}_1(t)$

第5章 結論

5.1 まとめ

本論文では、我々が聴きたい音を聴く際に行う“聞耳”を模擬する一手法として、目的音の音響的特徴を知識として利用する選択的分離抽出モデルを提案した。

まず、本論文では聴きたい音を聴くということを計算機上で実現するためには二つの課題があることを示した。一つ目は、我々が聴きたい音に注意を向ける行為を表現する問題であり、たくさんの音の中からどの音が目的音であるかを選択するという課題である。二つ目は、混合音中の目的音をどのように分離するかという問題であり、時間的にも周波数的にも重なりを持つような混合音から目的音成分のみ分離抽出するという課題である。

そこで、本論文ではこの二つの課題を同時に解決するために、楽器音の音響的特徴から得られるテンプレートを利用し目的音の選択を行うトップダウン処理と、目的音成分に重複した他音源成分の分離を行うボトムアップ処理を融合したモデルを構築した。

提案したモデルの有効性を示すために、様々な混合状況から目的音を分離抽出するシミュレーションを行った。まず、目的楽器音に雑音が付加した状況からのシミュレーションより、従来よく考えられてきた雑音中からの目的音の分離抽出に対しても本モデルが有効であることが確認できた。

次に、目的楽器音以外に複数の楽器音が混合された信号から目的楽器音を選択的に分離抽出するシミュレーションを行った結果、混合音中の目的音を正しく選択することが確認できた。さらに、目的楽器音成分の重複周波数成分分離についても、トップダウン処理、ボトムアップ処理単体での結果と比較し、分離精度の向上が確認できた。

最後に、三種類の楽器がそれぞれ異なる旋律を演奏する信号から、目的楽器の演奏のみを分離抽出するシミュレーションを行った。その際、楽譜情報に相当する新たな知識として、演奏を構成する単音の音高と、その発音順序をモデルに与えることとした。シミュレーションの結果、混合音中の目的楽器による演奏を選択的に分離抽出できることが確認された。これにより、本モデルは楽譜情報などの新たな知識を適用することで、楽器単音のみでなく楽器演奏音の分離にも応用できることを示した。

これらのシミュレーション結果より、本モデルが目的音の選択的な分離抽出に有効である事が示された。これにより、聴きたい音を聴き分けるという選択的な音源分離問題を考える場合に目的音に関する情報を利用することの重要性を示すことができた。

また、モデルに新たな知識表現を加えることで、さらに高度な問題へ発展できる可能性を示した。

5.2 今後の展望

以下に今後の展望、課題を示す。

1. 実際の音楽音響信号への適用に向けて

- 最適な知識表現の必要性

本論文では、目的音に関する情報として時間周波数平面上の形状で表現される特徴量テンプレートを利用した。今後、様々な楽器音を分離対象とするためには、それらのテンプレート作成が必要となる。しかし、現在のテンプレート生成部では、フルートとピアノの二種類のテンプレートが完成しているのみであり、その他の楽器については原信号を周波数分析して得られた瞬時振幅 $\tilde{A}_k(t)$ そのものをテンプレートとして利用していた。また、完成しているテンプレートについても改良の余地が残されており、各楽器音の特徴をさらに良く表現するものが望まれる。様々な楽器のテンプレートを効率良く作成する方法としては、数多くの楽器音サンプルを用いた最小二乗マッチングなどによる近似関数の最適化といった方法などが考えられる。

また、さらに高度な問題設定として実際の音楽への応用を考えた場合には、楽器単音の音響的特徴であるテンプレート以外にも知識表現が重要だと思われる。例えば、楽器演奏音の分離抽出シミュレーションで用いた楽譜情報のように単音同士のつながりに関する知識を利用することで、より実環境に近い状況からの分離抽出が可能であると考えられる。

- 基本周波数推定法の検討

本論文では、複数音源を含む音響信号からそれぞれの音源の基本周波数を推定する手法の構築を試みた。これにより、複数の音源が時間的に重複した場合でも、それぞれの音源の基本周波数を候補として抽出することが可能となった。ただし、この手法は「音源の基本周波数は持続時間中ほぼ一定である」という仮定に基づくものである。よって、この手法は本論文で扱ったような通常の演奏法による楽器音には適用できるが、スライド奏法や極端なビブラート奏法といった持続中に基本周波数が大きく変動する楽器音に対しては適用できない場合が生じる。これは、現状の手法が仮定を利用し、混合信号の持続時間全域にわたるヒストグラムから基本周波数候補を抽出しているため、時間的変化が考慮されていないということが原因である。

そこで、今後は現状のようなヒストグラムによる処理に代わって、図 3.5 のような時間周波数平面上の候補から連続的な基本周波数を抽出する処理を行うことで時間的変化に対応できると考えられる。

これによって、時間的に変動する基本周波数を抽出可能となれば、特殊な演奏法による楽器音と同様に、持続中に基本周波数が変動する人間の音声にも対応した手法となる可能性があると考えられる。

2. 音声など楽器音以外を対象とした分離抽出に向けて

本論文では、カクテルパーティ効果を模擬する一例として、分離抽出の対象を楽器音とした選択的分離抽出モデルの構築を試みた。しかし、第1章でも述べたように我々人間が持つカクテルパーティ効果という聴覚能力は、対象を楽器音や音声のみに限定するものではなく、様々な音事象に対してその効果を発揮する。ここで、音声などを分離対象とする場合には、本モデルがどのように応用できるかを考える。その一例として、混合音中から特定の音素を目的音として分離することが考えられる。その場合、目的音に関する特徴としてフォルマントなどのスペクトル情報をテンプレートに利用できると思われる。

また、話者認識などに利用される声紋などスペクトル形状の特徴を反映したテンプレートを実現できれば、特定話者音声を選択的に分離するといった応用も考えられる。この他にも、調音結合や文脈といった連続的な音声の特徴を考慮した知識表現を議論することで、音声を対象としたカクテルパーティ効果の実現に向けて発展できるものと考えられる。

謝辞

本研究を遂行するにあたり、数多くの御指導ならびに御鞭撻を賜りました赤木正人教授、党建武助教授に深く感謝の意を表します。また、日頃から熱心に御討論頂き、有益な御助言を賜りました、鵜木祐史助手に心より感謝致します。また、日頃から多大なる議論と激励を頂きました赤木研究室の諸先輩方に厚くお礼申し上げますと共に、本研究の遂行の多面に渡り御協力頂いた音情報処理学講座の皆様に感謝致します。

最後に、大学院での貴重な研究生活を与えて頂き、暖かく見守ってくれた両親、祖父母、兄弟、そして友人に心から感謝しお礼を申し上げます。

参考文献

- [1] Bregman, A.S., Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press, Cambridge, Mass., 1990.
- [2] Bregman, A.S., “ Auditory Scene Analysis: hearing in complex environments, ” in Thinking in Sounds, ed. S.McAdams and E.Bigand, pp.10-36, Oxford University Press, New York, 1993.
- [3] 鶴木祐史, 赤木正人, “ 聴覚の情景解析に基づいた雑音下の調波複合音の一抽出法 ”, 電子情報通信学会論文誌 Vol.J82-A No.10 pp.1497-1507 1999.
- [4] Masashi Unoki, Masato Akagi, “ A method of signal extraction from noisy signal based on auditory scene analysis, ” Speech Communication 27, pp.261-279, 1999.
- [5] 鶴木祐史 : 音の分離抽出における聴覚の計算理論に関する研究, 北陸先端科学技術大学院大学 博士論文, 1999.
- [6] 柏野邦夫, 中臺一博, 木下智義, 田中英彦 : “ 音楽情景分析の処理モデル OPTIMA における単音の認識 ”, 電子情報通信学会論文誌 Vol.J79-D-II, No.11, pp.1751-1761, 1996
- [7] 柏野邦夫, 中臺一博, 木下智義, 田中英彦 : “ 音楽情景分析の処理モデル OPTIMA における和音の認識 ”, 電子情報通信学会論文誌 Vol.J79-D-II, No.11, pp.1761-1770, 1996
- [8] Ellis, D. P. W. , Prediction-driven computational auditory scene analysis. Ph.D. thesis, MIT Media Lab, Massachusetts, 1996.
- [9] 中谷智広, 川端 豪, 奥野 博, “ 計算論的アプローチによる音響ストリームの分離, ” 音響学聴覚研究資, H-93-83, Dec. 1993.
- [10] 植田護, 橋本周司, “ 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, ” 情報処理学会論文誌, Vol.38, No.1, pp.146-157, 1997.
- [11] Unoki, M. and Akagi, M. “ Signal Extraction from Noisy signal based on Auditory Scene Analysis, ” In Proc. ICSLP98, Dec 1998.

- [12] M. P. Cooke, “Modeling Auditory Processing and Organisation,” Ph. D. thesis, CS dept. Univ. of Sheffield, 1991.
- [13] G. J. Brown, “Computational Auditory Scene Analysis: A Representational Approach,” Ph. D. thesis, CS-92-22, CS dept. Univ. of Sheffield, 1992.
- [14] 甘利俊一, “独立成分解析—脳データ解析の新しい手法,” Computer Today, 84, 1998.
- [15] 荒木章子, 牧野昭二, Robert Aichner, 西川剛樹, 猿渡洋, “死角型ビームフォーマを初期値に用いる時間領域ブラインド音源分離,” 音響学会講演論文集, pp.543-544, 2002.
- [16] 後藤真孝, “実世界の音楽音響信号を対象としたメロディーとベースの音高推定,” 情報処理学会 音楽情報科学研究会 研究報告 99-MUS-31-16, Vol.99, No.68, pp.91-98, 1999.
- [17] D.K.Mellinger, “Event Formation and Separation in Musical Sound,” Ph.D thesis, CCRMA, Standord Univ, 1991.
- [18] M.Mellody and G.H.Wakefield, “The time-frequency characteristics of violin vibrato:Model distribution analysis and synthesis,” J.Acoust. Soc. Am. Vol.107, No.1, 2000.
- [19] 櫻庭洋平, 河原達也, 奥野博, “定位情報と音色情報を用いた複数楽器音の認識,” 情報処理学会 音楽情報科学研究会 研究報告, No.46, 2002.
- [20] 後藤真孝: “音高による音色変化に着目した音源同定手法,” 情報処理学会 音楽情報科学研究会 研究報告 2001-MUS-40-2, Vol.2001, No.45, pp.7-14, May 2001.
- [21] 北原鉄朗, 後藤真孝, 奥野博, “音高による音色変化に着目した音源同定手法,” 情報処理学会 音楽情報科学研究会 研究報告, 2001-MUS-40-2, Vol.2001, No.45, pp.7-14, May 2001.
- [22] 木下智義, 坂井修一, 田中英彦, “周波数成分の重なり適応処理を用いた複数楽器音の音源同定処理,” 電子情報通信学会論文誌, Vol.J82-D-II
- [23] 柏野邦夫, 村瀬洋, “適応型混合テンプレートを用いた音源同定—複数楽器音への適用—,” 電子情報通信学会論文誌, Vol.J81-D-II, No.7, pp.1510-1517 1998.
- [24] 柏野邦夫, 村瀬洋, “動的メロディー抽出を用いたアンサンブル演奏の音源同定,” 音響学会研究誌, MA 97-4, pp.22-28, 199710-1517 1998.
- [25] 赤木正人, “カクテルパーティ効果とそのモデル化,” 電子情報通信学会誌解説, Vol.78, No.5 pp.450-453, 1995.

- [26] 赤木正人, “位相と知覚—人間ははたして位相聳か?—”, 日本音響学会平成8年度秋期研究発表会講演論文集, 1-2-2, 1996.
- [27] 赤木正人, “人間の聴覚特性を考慮した音声分離・強調,” 信学技報, SP2002-55, Vol.102, No.247, pp.25-30, 2002.
- [28] 石本祐一, 石塚健太郎, 相川清明, 赤木正人, “エントロピーによる重み付けを用いた雑音環境下での基本周波数推定,” 電子情報通信学会研究報告, Vol.102, No.247, pp.13-18, 2002.
- [29] 河原英紀, A. de Cheveigne “原理的に抽出誤りの存在しないピッチ抽出法とその評価について,” 信学技報, SP96 – 96, pp.9-18, 1997.

本研究に関する研究業績

- [1] 窪 正晃, 鷗木 祐史, 赤木正人, “ 楽器音の音響的特徴を知識として利用した選択的分離抽出法, ” 音響学会聴覚研究会資料, Vol. 32, N0.10, H-2002-90, Dec 2002.
- [2] 窪 正晃, 鷗木 祐史, 赤木正人, “ 複数音源からの目的楽器音の選択的分離抽出, ” 日本音響学会春季研究発表会, 1-5-15, March 2003