博 士 論 文

音の分離抽出における聴覚の計算理論に関する研究

指導教官　赤木 正人 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

鵜木 祐史

1999 年 1 月 14 日

目 次

記号の定義

記号	定義
$A_\ell(t), B_\ell(t)$	sinusoidal model における瞬時振幅（複合音）
$A_n(t), B_n(t)$	sinusoidal model における瞬時振幅（調波複合音）
ℓ, n	sinusoidal model における添字番号（整数値）
$\theta_{1\ell}(t), \theta_{2\ell}(t)$	sinusoidal model における瞬時位相（複合音）
ω_k, ω_ℓ, ω_n	中心角周波数
Ω_ℓ	sinusoidal model における帯域幅
$f_1(t)$	原信号（目的音）
$f_2(t)$	原信号（雑音あるいは目的音以外の音）
$f(t)$	混合信号
$X_{k,1}(t)$	分析フィルタにおける $f_1(t)$ の通過成分
$X_{k,2}(t)$	分析フィルタにおける $f_2(t)$ の通過成分
$X_k(t)$	分析フィルタにおける $f(t)$ の通過成分
$S_k(t)$	$X_k(t)$ における瞬時振幅
$\phi_k(t)$	$X_k(t)$ における瞬時位相
$A_k(t), B_k(t)$	各信号の $X_k(t)$ における瞬時振幅
$\theta_{1k}(t), \theta_{2k}(t)$	各信号の $X_k(t)$ における瞬時位相
$\theta_k(t)$	入力位相差
j	虚数単位
k	分析フィルタにおけるチャンネル番号（整数値）
$F_0(t)$	$f_1(t)$ の基本周波数
$C_{k,R}(t)$	$dA_k(t)/dt$ の R次区分多項式
$D_{k,R}(t)$	$d\theta_{1k}(t)/dt$ の R次区分多項式
$E_{0,R}(t)$	$dF_0(t)/dt$ の R次区分多項式
T_{S}	基本波の立上り
T_{E}	基本波の立下り
ΔT_{S}	高調波の立上り時間差
ΔT_{E}	高調波の立下り時間差
$T_{k,\mathrm{on}}$	$X_k(t)$ で検出された立上り
$T_{k,\mathrm{off}}$	$X_k(t)$ で検出された立下り

記号	定義
$A_k^{(R+1)}(t)$	$C_{k,R}(t)$ で Spline 補間された $A_k(t)$
$\theta_{1k}^{(R+1)}(t)$	$D_{k,R}(t)$ で Spline 補間された$\theta_{1k}(t)$
$\hat{f}_1(t)$	$f(t)$ からの $f_1(t)$ の分離抽出音
$\hat{f}_2(t)$	$f(t)$ からの $f_2(t)$ の分離抽出音
K	分析フィルタ数
R	多項式近似の次数
σ_A	$A_k(t)$ のなめらかさの尺度
σ_θ	$\theta_{1k}(t)$ のなめらかさの尺度
$\tilde{f}(a,b)$	wavelet 変換
a	スケールパラメータ
b	シフトパラメータ
α	スケール
D_ψ	許容条件の有限性を表す尺度
$\psi(t)$	アナライジング wavelet
$\hat{\psi}(\omega)$	アナライジング wavelet の Fourier 変換
$g_t(t)$	gammatone filter のインパルスの応答
$G_T(f)$	gammatone filter の周波数特性
a_f	gammatone filter のパラメータ（振幅）
N_f	gammatone filter のパラメータ（次数）
b_f	gammatone filter のパラメータ（帯域幅）
W	filterbank の解析可能な帯域幅
$\mathbf{Z}$	整数集合
$\mathrm{Comb}(\cdot)$	Comb filter
L_F	Comb filtering の周波数探索範囲の上限値
N_{F_0}	高調波の次数
t_h	基本周波数の不連続点
H	基本周波数の不連続点の個数
I	補間点
T_r	分離境界点
$\Delta\tau$	補間間隔
$F_{BP}(\cdot)$	帯域通過フィルタ

記号	定義
$\underline{C_{k,0}}$, $\overline{C_{k,0}}$	$C_{k,0}$の探索範囲の下限と上限
$\mathbf{y}_m$	観測信号
$\mathbf{x}_m$	状態変数
$\mathbf{v}_m$	観測雑音
$\mathbf{w}_m$	システム雑音
$\mathbf{F}_m$	状態遷移行列
$\mathbf{H}_m$	観測行列
$\mathbf{G}_m$	駆動行列

第 1 章

序論

1.1 はじめに

我々の聴覚には、特定の音を選択的に聞き取る能力（カクテルパーティー効果）や、瞬間的な妨害音によって欠落した情報を補って聞く能力（音韻修復現象）、いくつかの音の集まりをあるルールに従って一つの音の集まりにまとめて聞く能力（ストリーミング）といった様々な能力がある。これらは、聴覚が外界を理解するために備わった能力と考えられる。上記に示したような聴覚の優れた機能を、心理学、生理学、情報科学の立場から統一的に解明しようとする試みがある。これは、聴覚の計算理論の研究 [河原, 1994b] と呼ばれ、Marr によって提唱された視覚の計算理論 [Marr, 1982] の聴覚版に対応して、「聴覚が何を計算しているか、何故それをするのか」を説明するものである。

この計算理論を構築することは、もちろん聴覚機能の解明に直接継るわけであるが、それ以上に他分野に多大な貢献をもたらすことが予想される。例えば、工学的な側面として、電脳耳のようなコンピュータ上の聴覚モデルの実現だけでなく、これを利用した雑音にロバストな音声認識の実現や音源方向を推定するセンサなどにも応用できる。また、医療方面では、現存する補聴器に代わる、より人間の聴覚系に合致した補聴器に応用でき、心理学・生理学の分野では新たな知見を発見できる可能性もある。このように聴覚の計算理論の研究によるメリットは計り知れないものがあるが、計算理論を構築するための方法論が確立されていない。Marr による視覚の計算理論の概念が提案されてから、既に 15 年が過ぎようとしているが、聴覚の計算理論の研究に関しては、まだスタートしたばかりである。そこで、本研究では、聴覚の優れた機能の中でも、特定の音を分離抽出できるという機能に着目し、計算理論の構築の可能性を探る。

我々の身の回りの環境には、常に、様々な音源から発せられた音（話し声や雑音、残響、騒音など）が混在するわけであるが、聴覚はこのような環境の中でいともたやすく目的の音を分離抽出できる。では、数理工学的な立場から、同様の環境で、様々な音が混在する中から望みの信号を分離抽出するという問題を実現する場合はどうであろうか。残念ながら、この試みはとても困難な課題であり、実現には至っていない。しかし、この課題は、工学的応用の貢献が大きいことから、音声認識の研究だけでなく様々な信号処理の主要な研究課題になっており、数多くの方法が提案されている。

では何故、観測された混合信号から望みの信号を分離抽出するという課題が、数理工学的な処理として実現するには難しい問題であるのか。これは、混合信号から目的の音を分離抽出するという問題において、個々の音がどのように混合されたのかを表す情報が欠落していることに起因している。すなわち、この信号分離問題は、観測された信号から個々の信号を推定する不良設定の逆問題となっているからである。そのため、音や環境に対す

る何らかの制約条件がない限り、この問題を一意に解くことは難しい。

このように、数理工学の立場から、不良設定問題である音の分離抽出問題を一意に解くためには、音源や環境に対する制約条件が必要となる。しかし、我々の聴覚は無意識にこのような不良設定問題を容易に解いている。では、我々の聴覚は、どのような制約条件を用いて、どのような計算の方略で一意な解を求めているのだろうか。この答えを導く研究の一つが、聴覚の計算理論の研究であり、本論文の狙いである。

従って、本論文では、「二つの音を分離する」という基本的な機能に着目し、音の分離抽出における聴覚の計算理論の構築を試みる。

1.2 研究の背景

1.2.1 視覚の計算理論の研究

歴史的に見て、視覚の機能解明を狙いとした研究は、聴覚研究のものよりも古くから行われており、この研究として視覚の計算理論の研究がある。この研究は、Marr によって提案された概念が発端となり、多くの研究者によってその思想が引き継がれ、そして現在もなお発展し続けている。

Marr は、脳を理解するためには表 1.1に示すような三つの異なるレベルでの理解が必要であると主張した [Marr, 1982]。この三つのレベルでは、計算理論のレベル、表現とアルゴリズムのレベル、そしてハードウェアのレベルという順位で構成される。最上位にある計算理論は、計算の目標と計算の最適性、実行可能性を説いているのに対し、中位にあるアルゴリズムは、計算理論がどのように実現されるかを研究するものである。そして、最下位にあるハードウェアによる実現では、アルゴリズムを物理的にどうやって実現させるかという計算機構造を研究するものである。

Marr は、視覚という感覚器を切口に脳の機能解明を試みるため、視覚の計算理論の研究を行った。Marr による視覚の計算理論は、次の三つの主張で構成された。

1. 視覚の目的は、網膜上に投影された２次元の画像から外界の３次元の構造を推測することであるという主張

2. 視覚系は、モジュール構造（色、形、動き、陰影、テクスチャーなどを別々にする処理）をしているという主張

3. 初期視覚で計算された各視覚のモジュールの出力を全体として統合した２次元画像と３次元立体モデルの中間的な表現（$2\frac{1}{2}$次元スケッチ）が脳の中にあるという主張

表 1.1: 情報処理課題を実行する機械を理解するのに必要な三つの水準（Marr, 1982）

計算理論	計算の目標は何か、なぜそれが適切なのか、そしてその実行可能な方略の論理は何か。
表現とアルゴリズム	この計算理論はどのようにして実現することができるか。特に入力と出力の表現は何か、そして変換のためのアルゴリズムは何か。
ハードウェアによる実現	表現とアルゴリズムがどのようにして物理的に実現されるか。

また、Marr の視覚の計算理論では、画像データが３次元情報からデータが縮約され、落ちていることから、可視表面に関する何らかの拘束条件や事前知識がなければ初期視覚の問題は解けないことを指摘している [Marr, 1982]。ここで、Marr はある問題の (1) 解が存在し、(2) 解が一意であり、(3) 解が初期データに連続的に依存するとき、問題を良設定問題と呼び、この条件を一つでも満たさないとき、問題を不良設定であると呼んだ [Marr, 1982]。Marr の共同研究者であった Poggio は、この不良設定問題において、３次元物体から２次元画像への写像が光学の逆になっていることからこれを逆光学と呼び、標準正則化理論の枠組で不良設定問題を解くためのいくつかのアルゴリズムを統一できることを示した [Poggio *et al.*, 1985]。また、Geman and Geman は、確率場モデルを用いてこの画像復元問題を解いた [Geman and Geman, 1984]。そして、川人と乾は、これらの研究成果を発展させ、視覚大脳皮質の計算理論を提案した [川人, 乾, 1990]。

Marr が視覚の計算理論の概念を完成させるに当り、視覚の機能とは何であるのかという点で影響を受けた研究がある。これは、Gibson のアフォーダンス [Gibson, 1979] という考えである。視覚の計算理論の研究を計算的アプローチと呼ぶならば、Gibson の考え方は生態学的アプローチといえる。

1.2.2 生態学的視覚論の研究

アフォーダンスの概念は、対象の認知を可能にするのは形態自体ではなく、対象の変形の中にある不変項という特徴であり、環境が人にこの特徴を与える（アフォードする）というものであった [Gibson, 1979]。Gibson は、この概念を導く際、外界からの入力情報を処理したり、変形したり、推論したりする必要はなく、絶えず変化している感覚から、どのようにして日常生活における恒常的な知覚が得られるかという重要な問題を提起した。そして、感覚器は、外界の情報が生体にとってどんな意味があるかを「直接」に抽出できるように機能しているという考えに至った。つまり、この「外界が生体にとってどんな意

味があるか」という情報は、利用可能な形で外界に既に存在しており、視覚はそうした情報を抽出するようにつくられているということである。この概念は、生態光学と呼ばれる理論に発展し、現在では生態学的視覚論の研究に発展した。

Marr の視覚の計算理論の概念に影響を与えたのは、特に生態光学とアフォーダンスであると考えられる。また、Gibson の考えは、広く認知科学者らに受け入れられているが、Marr とは情報処理の立場から対極にあったように思われる。Marr 自身、Gibson の提起した問題の重要性や、視知覚の問題が外界の特性を感覚情報から復元することであるという点では思想的に合致していたようであるが、

- 物理的な不変項の検出は情報処理問題であること
- 不変項の検出をどのように実現するのか

という点で不十分であると批判していた [Marr, 1982]。しかし、上記の歴史的背景を概観すると、Marr と対極にあるような Gibson の研究があったからこそ、視覚の計算理論の概念が誕生したように思われる。

上記で述べた視覚研究における二つのアプローチは、聴覚研究にもそのまま当てはめることができる。次にこれらに対応する聴覚研究について説明する。

1.2.3 聴覚の計算理論の研究

視覚研究では、Marr による視覚の計算理論の概念が提案されて以降、急激に理論研究が展開されていったが、聴覚研究に至っては、ほとんど成されてこなかった。これには次のような大きな理由が考えられる。

一つは、視覚の計算理論の研究に比べ、聴覚の心理学的・生理学的知見が十分に得られていないことである。これは、生理学、解剖学の研究では、研究者の数が少なかったことも起因するかもしれないが、聴覚末梢系から奥の大脳皮質までの経路が複雑であったことや神経核自体が脳のかなり奥の方にあるため、その研究を困難にさせていた。

もう一つは、心理学、生理学、情報科学といった多分野にわたる聴覚研究者が共通の目的と方向性で研究を行っていなかったことも考えられる。最近、NATO ASI Computational Hearing [NATO ASI on Computational Hearing, 1998] における Cooke と Ellis による呼びかけ [Cooke and Ellis, 1998] や国内では河原による呼びかけ [河原, 1994b] により、聴覚の計算理論の研究が脚光を浴びつつある。

以上の理由により、Marr による視覚の計算理論のアナロジーから聴覚の計算理論を構築するためには、まだかなりの時間を必要とするものと思われる。

しかし、それでも現段階の聴覚心理学、生理学の知見を踏まえ、Marr の思想を引き継いで構築されてきた計算理論の研究がある。これには大きく分けて二つある。

一つは、Marr の後を継いだ MIT のグループによる研究である [Natural Computation, 1988]。このグループによって扱われた項目は、(1) 音とは何か、(2) 音響信号を表現する、(3) 両耳受聴による音源定位と音源分離の計算モデル、(4) 音声認識のためのスペクトログラムの大略化、(5) 歌声の音響学、(6) 音で材質を見分ける、(7) 壊れたのかバウンドしたのか？音で事象を判断することの心理物理学、(8) 旋律の知覚、であった。上記の報告は、確かに Marr の思想を引き継いでいるが、計算の本質を見究めるという意味では計算理論と呼ぶにはまだ不十分のように思われる。

もう一つは、入野によって提案された聴覚末梢系の計算理論である [入野, Patterson, 1994 ; 入野, 1995a]。これは Marr による初期視覚の計算理論に対応して議論されたものであり、gammachirp filter という聴覚フィルタ [Irino and Patterson, 1997] が、時間–スケール表現において最小不確定性の意味で最適であるというものである [入野, 1995a ; 入野, 1995b]。また、入野は聴覚心理現象を説明するための事象検出と強調を行うデルタガンマ理論についても述べている [入野, Patterson, 1994 ; 入野, 1995a]。この理論は聴覚末梢系と中枢系の一部の神経発火パターンを統一的な枠組で説明できるものである。入野の研究は、現在の計算理論の研究の中で、限りなく Marr の計算理論に近いものと思われる。しかし、この理論を確かなものにするためには、この理論を支持する、あるいは反例を示す聴覚の生理学的知見が出現するのを待たねばならないだろう。

さて、計算論的アプローチを計算理論の研究としたが、生態学的アプローチをとる研究は何になるだろうか。Gibson のアフォーダンスの概念はもちろん、感覚器を聴覚とした場合の研究に相当するが、Bregman によって提唱された聴覚の情景解析の研究もこれ対応するものと考えられる。この聴覚の情景解析の研究は、Gibson が生態光学を導く際に影響を受けたゲシュタルト心理学から得られたものである。

1.2.4 聴覚の情景解析の研究

Bregman は、自著の “Auditory Scene Analysis” で多くの聴覚心理実験の結果と考察をまとめた [Bregman, 1990]。これは、今までに個別に研究されてきた様々な聴覚心理現象を、音を通じて環境を把握するための機能として捉え直すことにより、統一的に理解しようとしたものである。ここで、情景解析とは、感覚から得られる証拠に基づいて外界の像を描き出すことを意味している。Bregman は、この情景解析の問題において、聴覚が利用している制約条件のいくつかを音響事象に関係する四つの心理学的な発見的規則：

(i) **共通の立上り/立下りに関する規則**

(ii) **漸近的変化に関する規則**

(iii) **調波関係に関する規則**

(iv) **一つの音響事象に生じる変化に関する規則**

として述べた [Bregman, 1993]。Bregman は、聴覚がこれら四つの発見的規則を用いて、受聴された音のかたまりを複数の音のストリームに分け、外界の解釈を試みている、と説明している。

さて、上記の解釈に基づけば、四つの発見的規則が「音とは何であるか」ということを表していることに気づく。つまり、四つの発見的規則の基づく思考は生態学的アプローチ以外の何者でもない。Gibson の主張した不変項では、Marr が指摘したようにどのようにそれを検出するかという点で取り扱いが難しい。しかし、Bregman によって提唱された発見的規則では、定量的な表現方法を除き、その取り扱いはさほど困難ではない。そのため、この発見的規則をどのように数学的に表現するかという点で、聴覚の情景解析の研究は、聴覚の計算理論を構築するための鍵になるかもしれない。

以上、視覚と聴覚の機能解明を試みる研究に対して、計算論的アプローチと生態学的アプローチをとる研究の背景をみてきた。次に、聴覚の計算理論の構築を直接の目的にしているわけではないが、とても深い関係にある三つの研究について述べる。

1.2.5 計算論的神経科学によるアプローチ

川人は、計算論的アプローチから脳機能を解明するために、次のように定義される計算論的神経科学を提案した [川人, 1996]。

> **脳の機能を、その機能を脳と同じ方法で実現できる計算機のプログラムあるいは人工的な機械を作れる程度に、深く本質的に理解することを目指すアプローチを計算論的神経科学と呼ぶ（川人, 1996）。**

川人による計算理論の構築方法は、Marr による視覚の計算理論の構築方法とは若干異なる。はじめに、計算理論を理解するために、その下位にある表現とアルゴリズムおよび神経回路モデルの研究を行う。次に、コンピュータシミュレーションや数値解析、ロボットを用いた制御実験、コンピュータビジョンの実験などを通し、妥当なモデルを実現することで計算理論を構築する。次に、心理実験、行動実験により構築された計算理論の検証を行う。これは異なる計算理論が構築されたときに、検証によって正しいものを絞り込む手

続きとなる。最後に、生理実験による神経回路モデルの検証を行う。この一連の方法論が、計算論的神経科学の本質であろう。

川人は、現在、心理学者、生理学者、数学者、工学研究者、行動学者らとチームを組み、上記の計算論的アプローチから

- 運動に関する研究

  眼球運動、手の到達運動、字を書く書字運動など

- 認知の情報処理に関する研究

  動きの検出、奥行きの知覚、３次元物体の面の向きの推定

といった研究に取り組んでいる。しかし、川人らの研究チームはまだ、聴覚の計算理論の構築に対する研究を進めていない。

1.2.6 数理工学的な信号分離の研究

聴覚の計算理論の構築を目的にしているわけではないが、数理工学的に信号分離問題を解こうとする試みは歴史的に古く、本論文ではすべてを紹介しきれないほど数多くの解法が提案されている。これまでに提案された方法を大別すると、音声認識の研究 [古井, 1985 ; Furui and Sondhi, 1991 ; 古井, 1995] では、背景雑音を除去あるいは抑圧する方法（例えば、[Boll, 1979] の研究）、音声そのものをエンハンスする方法（例えば、[Junqua and Haton, 1996] の研究）などがある。また、信号処理の分野では、入力信号と雑音を線形システムとして信号を推定する方法（例えば、[Shamsunder, 1997 ; Papoulis, 1977] の研究）や、信号や雑音を確率過程に従った信号として推定する方法（例えば、[Papoulis, 1991] の研究）などもある。最近では、音源情報から観測データを変換する順変換演算を、観測データから推測し、逆変換演算を求めることで音源を分離する Blind Separation という方法もある [Shamsunder, 1997]。

これらの方法は、不良設定の逆問題を解くという意味で考えれば、音源情報の独立性（無相関）、確率的な独立性の仮定、雑音が白色ガウス過程に従うといった様々な拘束条件を設けることで一意な解あるいは近似解を導いていると解釈できる。しかし、この研究を聴覚の計算理論の研究に発展させるためには、制約条件の意味づけや音を分離する際の「入力、出力、処理過程」を明らかにしなければならない。

1.2.7　計算論的な聴覚の情景解析の研究

計算論的な聴覚の情景解析（CASA: Computational Auditory Scene Analysis）という研究がある。これは、聴覚の情景解析の問題を計算モデルとして実現する試みである。現在、主に研究されている課題は、音源分離問題を聴覚の情景解析問題としてとらえ直し、それを計算機上に実装することである。この計算モデル（表 1.2参照）には、大きく分けてボトムアップ処理に基づくモデルとトップダウン処理に基づくものがある。前者の代表的なモデルとして、音響事象に基づいた分離モデル [Cooke, 1991 ; Brown, 1992 ; Cooke and Brown, 1993 ; Brown and Cooke, 1994] や、基本周波数の推定に基づいた二重母音の分離モデル [de Cheveigné, 1993 ; de Cheveigné, 1997]、楽音の分離モデル [柏野, 田中, 1994a ; 柏野, 1994b] がある。ここで、Cooke や Brown による分凝モデルを例にとってみよう。このモデルは、聴覚末梢系、聴覚マップ表現、音源分離という三つの処理過程で構成される。モデルで観測された混合信号は、gammatone filterbank と神経発火モデルから成る聴覚末梢系により、(1) 神経発火率、(2) 周波数遷移、(3) 立上り、(4) 立下り、(5) 自己相関、(6) フィルタ間の相関といった抽象表現（computational map と呼ばれる）と、強度、強度差、スペクトル形状、同期性に表現される。そして、これらの抽象表現を基に、聴覚的な要素が形成され、最後に情景解析モデルにより、各要素を同じ音源で生じる成分にグルーピングし、再構成することで音源分離を実現している。

次に後者の代表的なモデルとして、心理音響学的なグルーピング規則に基づいたモデル [Ellis, 1996] やマルチーエージェントシステムによるストリーム分離モデル [中谷ら, 1993 ; Nakatani *et al.*, 1994 ; Nakatani *et al.*, 1995a ; Nakatani *et al.*, 1995b ; 中谷ら, 1995] がある。ここで、Ellis による分凝モデルを例にとってみよう。この分凝モデルは、人間の聴覚系で行われている体制化と分凝の有効なシミュレーション結果を得るためのプロトタイプであり、心理学的知見を重視したものである。このモデルは、(1) 音生成事象、(2) 時間–周波数解析（蝸牛殻フィルタバンク）、(3) 初期グルーピング、(4) ２次グルーピング、(5) 高次処理、で構成される。また、モデル内の音事象は、蝸牛殻の機能を模擬したフィルタバンクにより時間–周波数表現に解析され、初期グルーピングでは、部分的なオブジェクトを生成する。この初期グルーピングは、調波性、共通な立上り、近接、連続性といった基本的な心理音響学的手がかりを識別するための規則で構成されている。２次グルーピングでは、部分オブジェクトから基本オブジェクトを生成する。この２次グルーピングは、初期グルーピングで生成された沢山の候補を絞り込むために、切り取り、相関、近接といった規則で構成される。そして最後に、高次処理により、分凝に必要な知識を得ることで音源分離を実現している。

表 1.2: CASA の研究で提案された音源分離の計算モデル

モデル	特徴
音響イベント（基本周波数、立上り/立下り、AM など）に基づいた分凝モデル [Cooke, 1991 ; Brown, 1992]	ボトムアップ
心理音響学的グルーピングの規則に基づいた分凝モデル [Ellis, 1994]	トップダウン
予測駆動型の分凝モデル [Ellis, 1996]	トップダウン
ベイジアンネットワークによる情報統合の機能を備えた音楽情景分析の処理モデル：OPTIMA [柏野ら, 1996a ; 柏野ら, 1996b]	ボトムアップ /トップダウン
基本周波数推定とキャンセレーションモデルを用いた二重母音の分離モデル [de Cheveigné, 1993]	ボトムアップ
マルチエージェントシステムによるストリーム分凝 [中谷ら, 1993]	トップダウン
聖徳太子コンピュータ [Okuno *et al.*, 1997]	トップダウン
ラウドネス・ピッチ・音色に基づいた分凝モデル [Abe and Ando, 1997]	ボトムアップ
スペクトル形状変化の追跡・予測による音声分離 [Katsuse *et al.*, 1997]	ボトムアップ

以上、二つのモデルを簡単に紹介した。これらのモデルのほとんどは、Bregman によって提唱された発見的規則 (i) と (iii) を利用したものであり、音響的な特徴として振幅（あるいはパワー）スペクトルを用いている。そのため、工学的な問題ではあるが、これらのモデルでは、二つの信号が同じ周波数領域の成分を含むような場合、二つの信号を完全に分離できているとは言い難い。また、計算理論の研究という観点では、これらの研究には、

- アルゴリズムの研究から計算理論の研究に発展させる構築方法の構想が明確でない。
- アルゴリズムで利用する発見的規則に対して、数学的な議論（十分性や必要性など）がほとんど行われておらず、単なる実装に留まっている。
- 様々なアルゴリズムを統一的に検証しようとする試みがない。

という問題点がある。これらのうち、ひとつでも克服できない限り、アルゴリズムの研究から計算理論の研究への発展は望めない。

最近になって、Cooke と Ellis はこれまでの CASA で取り上げられてきた知覚現象とその計算モデルを対比させ、本質的な計算の目的を追求することで、CASA 研究を計算理論の研究に発展させる試みを始めている [Cooke and Ellis, 1998]。彼らは、特に音のグルーピングに着目し、聴覚的体制化の計算理論を統一的に確立しようと試みている。

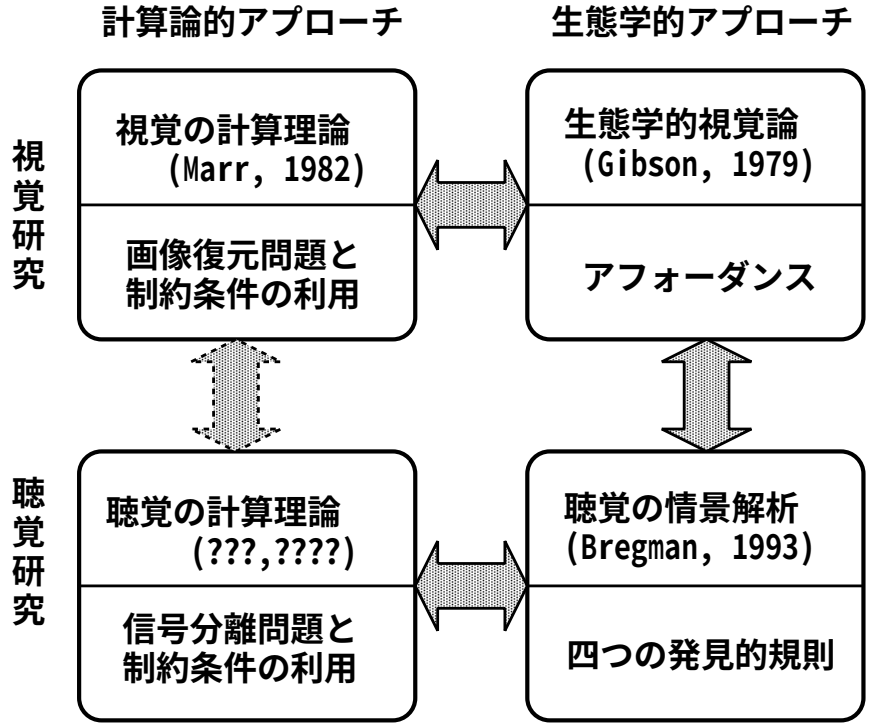

図 1.1: 計算論的アプローチと生態学的アプローチ

1.3 本研究のアプローチ

上記の研究背景でみたように、聴覚の心理学的知見と生理学的知見が十分に得られていないため、現段階では、視覚の計算理論のアナロジーとした聴覚の計算理論を構築するためには、まだかなりの時間を必要とする。しかし、十分な知見がそろうまで研究を進めないということは科学の進歩を止めるのに等しいと思える。おそらく別の観点から計算理論の構築を目指す方法があるはずである。

例えば、数理工学的なアプローチに固執した場合では、解法に利用された制約条件の意味づけが不十分であるが、心理学的にあるいは生理学的に意味のあるものであれば計算理論の研究に発展するだろう。また、計算論的な聴覚の情景解析のアプローチを取る研究では、発見的規則（制約条件）の実装に留まり、アルゴリズムの研究の域を出ていないが、先のものとは反対に数理工学的な意味づけをすることができれば計算理論の研究に発展できる。つまり、不良設定問題を一意に解くために利用する制約条件に対し、心理学的、生理学的、情報科学的意味での必要性、最適性などを議論することができれば、それは計算理論の研究と成り得る。

そこで、先の研究背景について、不良設定問題に対する解釈を再考してみる。Marr の視覚の計算理論の研究に始まり、以後続いている計算理論の研究の一連の流れには、「欠落し

た情報から元の情報を推定する情報復元問題を解くためには制約条件が必要である」という統一的な考え方があった。そこで、混合信号から目的の音を分離抽出するという聴覚の情景解析問題を解くためには制約条件が必要であるという考えを本研究でも採用すれば、音の分離抽出における聴覚の計算理論を構築できる可能性がある。

次に、Marr が視覚の計算理論の概念を完成させるに当り、Gibson の思想に影響を受けたように、聴覚の計算理論を構築する際にも生態学的アプローチをとる聴覚の情景解析の研究に大いに影響を受けるものと思われる。このヒントは、聴覚の情景解析の研究が与えてくれる。つまり、Bregman によって提唱された発見的規則に対して、数学的に考察（十分性・必要性など）することが計算理論の構築につながるということである。従って、発見的規則は、定性的な条件であるが故に、数理工学的な制約条件として直接利用できないものの、聴覚の情景解析という名の不良設定問題を一意に解くための制約条件として利用できれば、音の分離抽出における聴覚の計算理論を構築できる可能性がある。

1.4 本論文の目的

前節に示した考え方から、本研究では図 1.1に示すように、

- 混合された信号から目的の原信号を求める信号分離問題を一意に解くためには、音や環境に対する制約条件が必要である。
- 信号分離問題を聴覚の情景解析問題としてとらえ直せば、信号分離問題を一意に解くために必要な制約条件として聴覚が情景解析問題で利用している心理学的な制約条件を利用できる。

という立場を取る。そして、計算論的神経科学のアプローチで利用された方法論と同様、アルゴリズムの研究から計算理論の研究へと発展させることで、音の分離抽出における聴覚の計算理論の構築を試みる。従って、本研究では、聴覚系でどのような制約条件を設けることで目的の処理（不良設定問題を一意に解くこと）が可能なのかを、アルゴリズムより一段上のレベルで検討しなければならない。また、この制約条件を利用したアルゴリズムが計算理論から導かれたものになるためには、制約条件の必要十分性を示さなければならない。本研究では、心理学的・生理学的に意味のある制約条件を利用したアルゴリズムであり、かつ上記のようにアルゴリズムの研究から計算理論の研究へと発展する道筋が明確にあるアルゴリズムを「計算の方略」と呼ぶ。ここで、計算の方略は、Marr の示した計算理論でいう「何を計算しているのか」を説明できるものであるが、その必要十分性が導かれることにより、「何故それをするのか？(計算の目的)」も説明できることになる。しか

し、この必要十分性を導くためには、沢山の計算の方略を提案し、聴覚心理実験・生理実験によりこれらを検証することで、正しいものに絞り込まなければならないため、多くの時間を必要とする。

そこで、本論文では、「二つの音を分離する」という基本的な聴覚の機能に着目し、不良設定問題である信号分離問題を、二波形が加算されたものから個々の波形に分離抽出するという「二波形分離問題」に限定する。そして、計算理論を構築するための一歩として、「どのような制約条件を用いることで二波形分離問題を一意に解くことができるか」という戦略的な解法、つまり妥当と思われる聴覚の計算の方略を明らかにする。

はじめに、音の振幅スペクトルと位相に着目した二波形分離問題の定式化を行い、Bregman によって提唱された四つの発見的規則を定式化する。その後で定式化された制約条件を利用することで二波形分離問題の解法を導出する。次に、分離抽出の対象となる信号を AM–FM 調波複合音と定義し、単純な音（正弦波）からより複雑な音（調波複合音）に発展させた分離抽出音を利用して、二波形分離問題で利用する制約条件の十分性を検証する。最後に、AM–FM 調波複合音を利用して二波形分離問題音で利用する制約条件を順次省略した場合の分離精度を評価することで、制約条件の有効性を示す。そして、二波形分離問題において必要な物理量と制約条件を議論することで、音の分離抽出における聴覚の計算の方略を導出する。

上記の方法論で、音の分離抽出における聴覚の計算の方略を構築することができれば、共変調マスキング解除といった聴覚心理現象のモデル化や、工学的側面として、雑音に頑健な音声認識システムのフロントエンドとしての応用に期待できる。また、計算の方略で利用される制約条件の必要十分性を導くことができれば、音の分離抽出における計算理論を構築することができる。さらに、この計算理論は、カクテルパーティ効果のモデル化の手がかりとして聴覚の情景解析の研究に貢献するだけでなく、視覚の計算理論のアナロジーとして聴覚の計算理論の構築を試みる研究に対しての方向性も提供できる。

1.5　本論文の構成

本論文は７章で構成される。

第１章では、本論文が対象としている研究分野の背景と問題点を指摘し、本論文の位置付けと目的を示す。

第２章では、音の分離抽出における聴覚の計算の方略を構築するための方法論を提案する。はじめに、本論文で取り扱う信号分離問題として二波形分離問題の枠組を示す。次に、本論文で取り扱う信号音を AM–FM 複合音の物理的表現で定義し、四つの発見的規則を再

考することで分離抽出の対象となる音を仮定する。その後で、仮定した分離抽出の対象となる音の網羅性を議論し、二波形分離問題で利用する制約条件の概念を述べる。最後に、計算の方略を構築するための方法論として、発展的構築法を提案する。

第３章では、二波形分離問題の解法を提案する。はじめに、AM–FM 複合音で定義される信号分離問題として二波形分離問題を定式化し、この問題が不良設定の逆問題であることを示す。次に、この不良設定問題を一意に解くために発見的規則を制約条件に定式化する。そして、二波形分離問題の解法を実現するモデルを実装する。最後に、分析フィルタ群や音の物理量の計算方法を示し、AM–FM 調波複合音を分離抽出できることを示す。ここでは、原信号と分離抽出された信号の差異が定量的に少ないとき、つまり妨害音が取り除かれ、定量的に差異が改善されたときに、二波形分離問題で利用された制約条件が十分性を満たしたものと定義する。例えば、原信号に 20 dB の雑音が付加されたとき、分離抽出された信号と原信号の SNR が 25 dB であったとする。このとき、分離抽出された信号は、5 dB の分離効果が見られる。従って、本論文では、このように分離効果が見られたときに、二波形分離問題で利用した制約条件の十分性が示されたとする。

第４章では、AM–FM 調波複合音を利用して、第３章で提案した二波形分離問題の解法の十分性と有効性を検証することを目的とする。はじめに、AM 単一成分音を利用して、二波形分離問題の解法の十分性を検証する。この検証では、単一成分における瞬時振幅に対して、漸近的変化の制約条件の十分性が検証される。次に、AM–FM 調波複合音を利用して、二波形分離問題の解法の十分性を検証する。この検証では、複合成分における瞬時振幅および瞬時位相に対して、漸近的変化の制約条件の十分性が検証される。また、これは基本周波数の時間変動の有無による影響も検証できる。最後に、AM–FM 調波複合音を利用し、二波形分離モデルで利用する制約条件を順次省略した場合の分離精度を評価することで、制約条件の有効性を検証する。

第５章では、音の分離抽出における計算の方略を提案する。これは、検証後に得られた二波形分離問題の解法から、音の分離抽出に必要な物理量、物理的表現、制約条件について再考することで導出される。ここでは、分離抽出の対象となる信号を AM 調波複合音と仮定し、制約条件を十分条件とした聴覚の処理機能の「入力、出力、処理過程」を導く。

第６章では、実音声を対象とした二波形分離問題と共変調マスキング解除を想定した二波形分離問題に対し、本論文で提案した計算の方略を展開することで、この計算の方略がこれらの問題の解法を導出できることを示す。

第７章では、本論文で得られた結果を要約し、今後の展望を述べる。

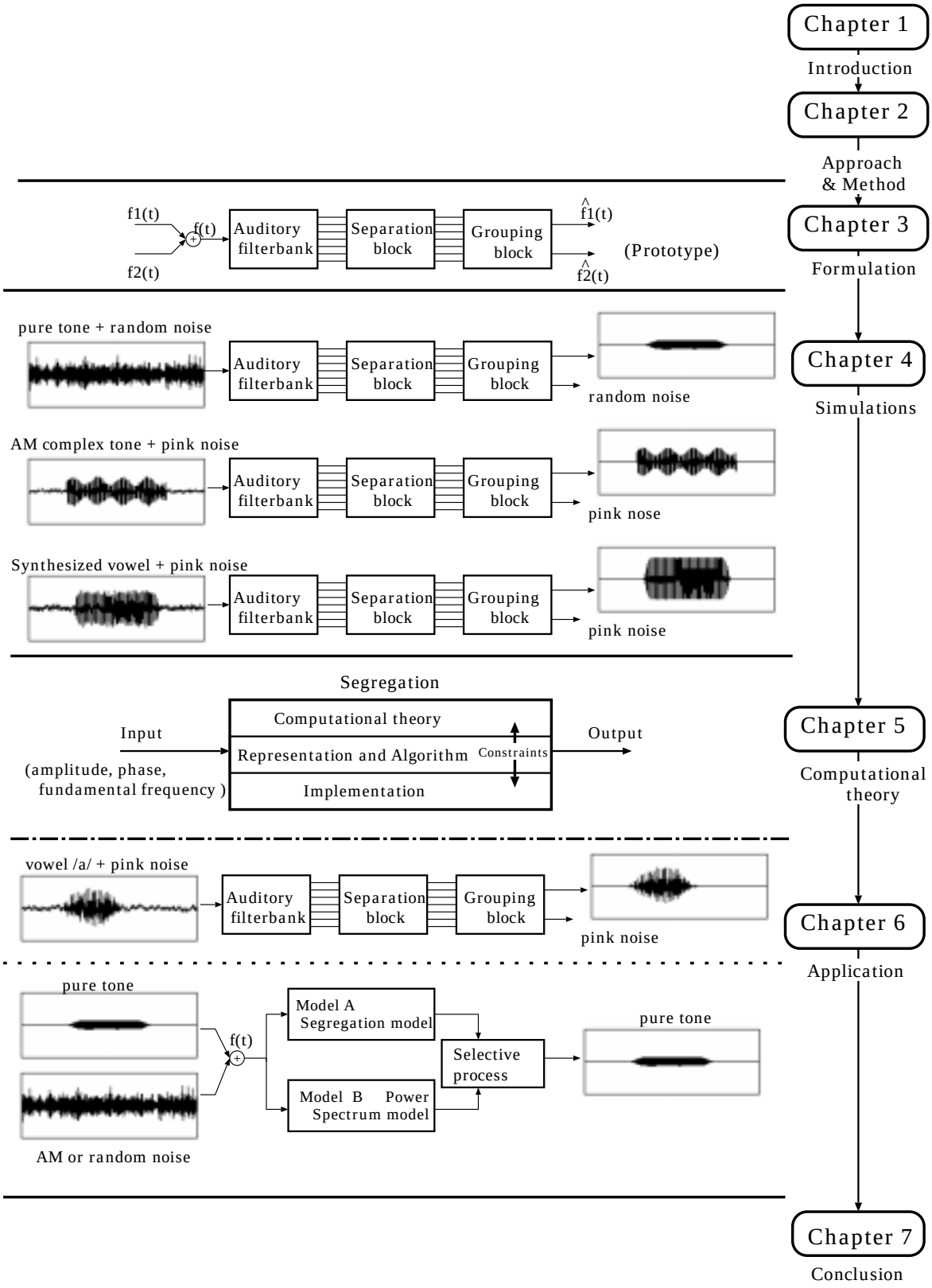
Chapter 1
Introduction
Chapter 2
Approach & Method
Chapter 3
Formulation
Chapter 4
Simulations
Chapter 5
Computational theory
Chapter 6
Application
Chapter 7
Conclusion
f1(t)
f2(t)
f(t)
Auditory filterbank
Separation block
Grouping block
(Prototype)
pure tone + random noise
random noise
AM complex tone + pink noise
pink nose
Synthesized vowel + pink noise
pink noise
Segregation
Input
(amplitude, phase, fundamental frequency)
Computational theory
Representation and Algorithm
Implementation
Constraints
Output
vowel /a/ + pink noise
pink noise
pure tone
Model A Segregation model
Model B Power Spectrum model
Selective process
AM or random noise

図 1.2: 論文の構成

第 2 章

聴覚の計算の方略を構築するための方法論

2.1 まえがき

本章は、音の分離抽出における聴覚の計算の方略を構築するための方法論を提案することを目的とする。

はじめに、「二つの音を分離する」という基本的な聴覚の機能に着目し、本論文で取り扱う信号分離問題を、二波形が加算されたものから個々の音に分離抽出するという二波形分離問題とする。

次に、本論文で取り扱う信号音を sinusoidal model として知られる AM–FM 複合音の物理的表現で定義する。その後で Bregman によって提唱された四つの発見的規則を考察することで、二波形分離の対象となる音を定義する。また、この分離抽出の対象となる音の網羅性を議論し、二波形分離問題を解くために利用する制約条件の概念を述べる。

最後に、音の分離抽出における聴覚の計算の方略を構築するための方法論として、発展的構築法を提案する。この方法では、AM–FM 調波複合音を利用することで二波形分離問題の解法の十分性と有効性を検証することで計算の方略を導出する。この構築法の名前は、最も単純な音（AM 単一成分音）からより複雑な音（AM–FM 調波複合音）へと発展的に問題を拡張して検証することに由来してつけられた。

2.2 信号分離問題の枠組

我々の身の回りの環境には、常に、様々な音源から発せられた音（話し声や雑音、残響、騒音など）が混在するわけだが、聴覚はこのような環境の中でいともたやすく目的の音を分離抽出できる。例えば、日本には、人間の優れた聴覚の機能を示す一つの伝説がある。その伝説とは、「聖徳太子が 10 人の訴えを同時に聞き分け、それを処理した」といわれていることである。我々一般人には、聖徳太子と同様の処理を行うことは難しいと思われるが、例えば誰か一人の声を聞き取るということはとても容易な処理であろう。このような様々な音が混在する中からある特定の音を選択的に聞き取る、すなわち、分離抽出する処理は、カクテルパーティー効果と呼ばれている [Cherry, 1953]。この名前は、とても賑やかなパーティー会場の中でも、特定の話者の声を聞き取ることができる、ということに由来する。カクテルパーティー効果が起こる原因は完全に解明されたわけではないが、音の到達方向の違い、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ものと考えられている [赤木, 1995]。

さて、図 2.1に示すように、カクテルパーティー効果には、知覚メカニズムとして二つの問題が介在している。一つは、Speaker A〜Z の音の中からどの音を取り出すかという問題

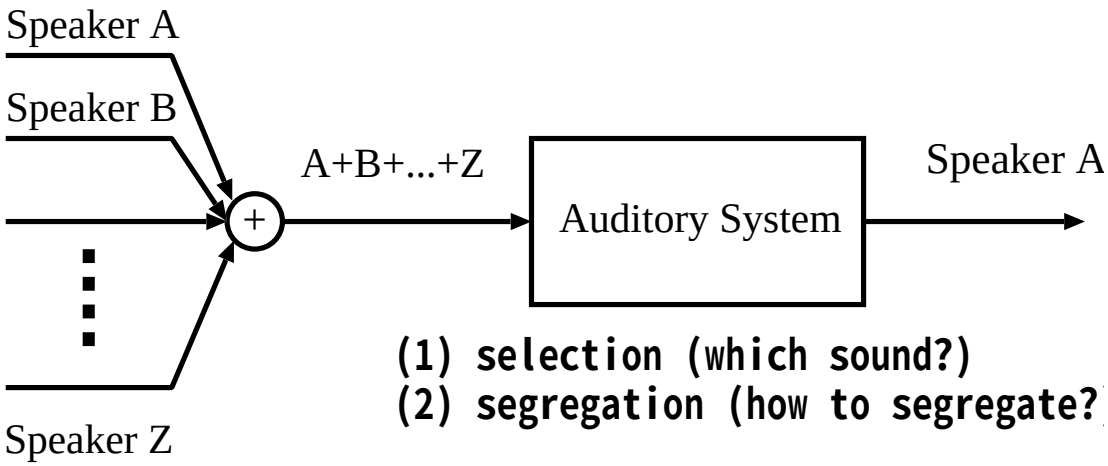

図 2.1: 信号分離問題の枠組

であり、もう一つはそれらをどのようにして分離抽出するか（例えば、A〜Z の中から A をどのようにして分離抽出するか）という問題である。前者の問題は、音に対する注意の研究課題であり、後者の問題は信号分離の研究課題である。本研究では「二つの音を分離する」という基本的な聴覚の機能に主眼をおいているのだから、当然、後者の問題が主問題となる。そのため、本論文ではカクテルパーティー効果のような一般的な知覚上の分離抽出問題から、注意の問題を取り除いて考えなければならない。また、Speaker を A〜Z ではなく、A と B に限定した基本的な信号分離問題を想定しなければならない。

そこで本章では、不良設定問題である信号分離問題を、“二つの信号だけが存在し、それらが混合した状態からそれぞれの信号を分離する” という問題とする。本論文では、この問題を二波形分離問題と定義する。そして、聴覚の計算の方略を導くために、どのような制約条件を設けることで二波形分離問題を一意に解くことが可能なのかを検討する。

2.3 分離抽出音の定義

2.3.1 信号音の物理的表現

前節では、信号分離問題を二波形分離問題と定義したが、ここでは本論文で取り扱う音をどのように物理的に表現するかを議論する。我々の身の回りには、音声や楽器音、あるいは金属音など様々な音が存在する。また、これらの音は何かの物体に当りその反射音としても存在する。これらの音を物理的に表現するためにはどうしたらよいだろうか。

はじめに、最も単純な音として純音を考える。純音は、単一周波数成分をもつ正弦波であ

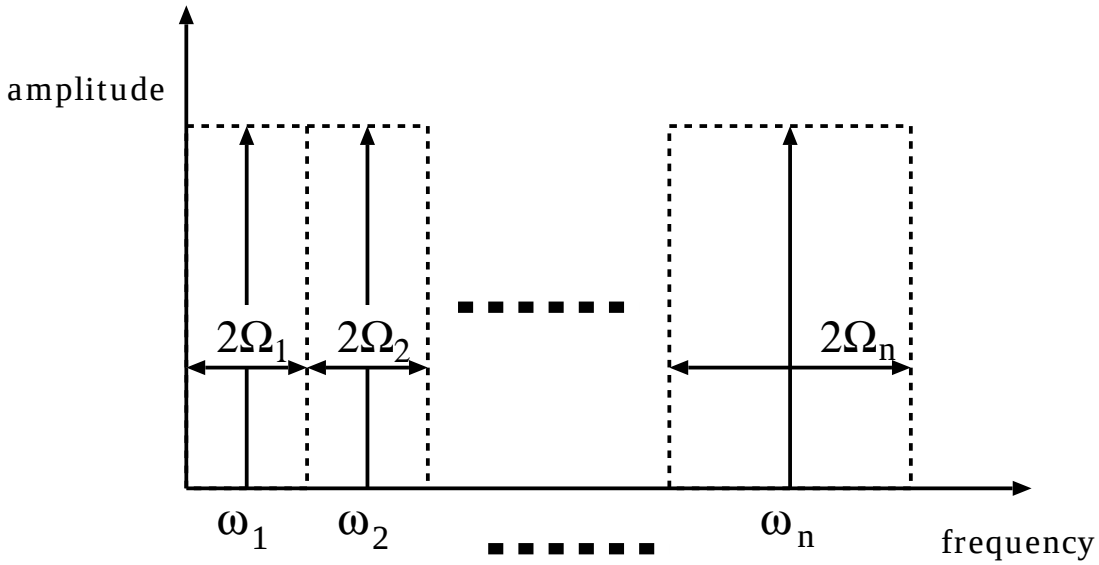

図 2.2: AM–FM 複合音の周波数成分

る。そこで、次にこの純音に対し、時々刻々と振幅が変化する状態を考える。これは正弦波の振幅が時間的に変化することを意味するため、振幅変調（AM: Amplitude Modulation）の形式として表現する。同様に、音の単一周波数も時々刻々と変化する場合を考える。これは、正弦波の周波数も時間的に変化することを意味するため、周波数変調（FM: Frequency Modulation）の形式として表現する。最後に、音は単一成分音の集まりであるから、これを AM–FM 複合音として表現できる。そこで、図 2.1にあるように二つの音源で生じた二つの信号を、それぞれ $f_1(t)$ と $f_2(t)$ とおき、

$$f_1(t) \approx \sum_{\ell} A_\ell(t) \exp(j\omega_\ell t + j\theta_{1\ell}(t)) \tag{2.1}$$

$$f_2(t) \approx \sum_{\ell} B_\ell(t) \exp(j\omega_\ell t + j\theta_{2\ell}(t)) \tag{2.2}$$

と近似的に表現する。但し、$A_\ell(t)$ と $B_\ell(t)$ は振幅変調項を表す瞬時振幅、$\theta_{1\ell}(t)$ と$\theta_{2\ell}(t)$ は位相変調項を表す瞬時位相である。ここで、周波数変調項は位相変調の形式で表現される。また、ℓは整数値、ω_ℓは中心角周波数であり、自由な値を選べるものとする。

これは、sinusoidal model [McAulay and Quatieri, 1986] として知られる表現でもあり、一般にすべての周波数帯域を網羅し、その範囲内で ω_ℓを自由に設定することですべての音を表現することができる。何故かというと、図 2.2に示すように ω_ℓの値が離散的であったとしても隣合う ω_ℓ間の周波数差を $A_\ell(t)$ に含まれている最大周波数 Ω_ℓと考えれば、このような隙間をすべて埋めることが可能だからである。同様に、これは $\theta_{1\ell}(t)$ でも補うことができる。従って、我々の身の回りに存在する音は理論的に上記の表現で記述できる。しか

し、実際上は無限周波数まで利用することは不可能であるから、帯域制限された範囲内でのすべての音を表現できることになる。

次に、上記の音の物理的な表現形式に基づき、分離抽出の対象となる音を定義するため、Bregman によって提唱された四つの発見的規則を本論文でも考察する。

2.3.2 四つの発見的規則の解釈

Bregman は、聴覚が情景解析問題を解くために利用している制約条件のいくつかを音響事象に関係する四つの心理学的な発見的規則として述べている [Bregman, 1993]。

(i) **共通の立上り/立下りに関する規則**

この発見的規則は、“無関係な音は同時に始まったり終ったりしない” ということを意味する。言い換えると、“ある一つの音源で生じた音は、同時に始まったり終ったりするが、別の音源で生じた音は前者の音と同時に始まったり終ったりしない” ということを意味している。

(ii) **漸近的変化に関する規則**

(a) 単一の音はその性質上、ゆっくりと滑らかに変化する傾向がある。

(b) 同じ音源で生じる一連の音は、その性質上ゆっくりと変化する傾向がある。

この発見的規則は、“同じ音源で生じた音は急激に変化しない” ということを意味する。言い換えると、急激な変化が生じた場合に、別の音源で生じた音を含んでいるということを意味している。また、漸近的変化は、時間軸上と周波数軸上の両方で考えられる。

(iii) **調波関係に関する規則**

この発見的規則は、“繰り返し周期で一連の音が振動するとき、その変動は周波数成分が共通な基本周波数の整数倍になるように聴覚パターンを発生する” ということを意味する。

(iv) **一つの音響事象に生じる変化に関する規則**

この発見的規則は、“一つの音響事象で生じる多くの変化は、同時に同じように音の成分すべてに影響を与える” ということを意味する。

Bregman は、聴覚がこれら四つの発見的規則を用いて、受聴された音のかたまりを複数の音のストリームに分け、外界の解釈を試みている、と説明している。例えば、聴覚心理

物理実験から、音の立上りや立下りの周波数間での同期性 [柏野, 田中, 1994a]、基本周波数に対する倍音構造の有無やずれ [Darwin *et al.*, 1992 ; Darwin *et al.*, 1994]、変調のコヒーレンス性（振幅変調や周波数変調）[Bacon, 1989] という物理的な特徴が、音のかたまりを二つのストリームに分凝する、ということが報告されている。また、聴覚的誘導や音韻修復といった錯覚と考えられている聴覚心理現象は、発見的規則 (i) と (ii) を主に用いて聴覚系が信号を能動的に解釈した結果であると考えられている [河原, 1994a]。一方、共変調マスキング解除 [Hall and Fernandes, 1984] も、発見的規則 (iv) を利用して妨害音を能動的に分離する情景解析の機能であると考えられている [Bregman, 1993]。尚、上記の四つの発見的規則は、様々な聴覚心理実験の結果から得られたものであり、この四つが制約条件のすべてである保証はない。また、四つの制約条件すべてを必ず利用しているとも限らない。

以上のように、これらの発見的規則は、音とはどのようなものなのかという Bregman の思想を述べているものと解釈できる。つまり、この思考は音や環境に対する生態学的アプローチ以外の何者でもない。このことから、四つの発見的規則のいずれかを満たすものは、我々の環境の中にある一つの音源で生じた音であると解釈できる。しかし、四つの発見的規則は実に曖昧な表現であり、あらゆる音に対しても解釈の仕方によっては適用できたり、できなかったりする。また、先に述べたように同時マスキングや共変調マスキングといった様々な状況によっても四つの発見的規則のうちのいくつかを利用したり、あるいはすべてを使ったりもする。このような曖昧な発見的規則を厳密な（数理工学的な）制約条件として定式化するためには、分離抽出音を厳密に定義し、その上でこれを拘束する制約条件を議論する必要がある。

そこで、四つの発見的規則すべてを利用する状況をはじめに想定する。その後で、Bregman が四つの発見的規則を見い出したように、本論文でも Bregman が述べるような拘束条件を受ける音とはどのようなものなのかを議論する。

2.3.3 分離抽出の対象となる音の仮定

はじめに、発見的規則 (i) を考える。これは、一つの音源で生じた音は同時に始まって、同時に終ることを意味するのだから、音を構成する各周波数成分も同時に始まり、同時に終るものと解釈する。従って、式 (2.1) から、この発見的規則によって拘束される音とは、複合音を構成する各単一成分音の始まりと終りが同期している複合音にほかならない。そのため、立上りと立下りの情報は、各瞬時振幅 $A_\ell(t)$ に現れる。

次に、発見的規則 (ii) を考える。これは、一つの音源で生じた音は時間的にも周波数的に

も急激に変化しないのだから、時間的にも周波数的にも不連続点をもたないものと解釈する。つまり、物理的な音としては、時間的にも周波数的にも連続でなめらかであるような音の物理量をもつということである。従って、式 (2.1) から、この発見的規則によって拘束される音とは、各瞬時振幅 $A_\ell(t)$ と瞬時位相 $\theta_{1\ell}(t)$ が連続的に変化する音にほかならない。

次に、発見的規則 (iii) を考える。これは、一つの音源で生じた音が調波構造を持つことを意味するのだから、その音の基本周波数の整数倍に高調波が存在すると解釈する。従って、式 (2.1) から、この発見的規則によって拘束される音とは、各単一周波数成分音の重ね合わせ方が基本周波数の整数倍になるような複合音にほかならない。そのため、この調波構造の情報は、$\omega_\ell/2\pi$のとる値に現れる。ここで、この規則に対応する物理的な音は、有声音（母音）や和音といった調波複合音である。

最後に、発見的規則 (iv) を考える。これは、一つの音源で生じた音であれば、振幅包絡の変動が一致することを意味するのだから、複合音が共通の信号で振幅変調されるものと解釈する。従って、式 (2.1) から、この発見的規則によって拘束される音とは、共通の振幅変調を受ける AM–FM 複合音にほかならない。そのため、この情報は各瞬時振幅 $A_\ell(t)$ のコヒーレンス性に現れる。

以上の考察を踏まえた結果、四つの発見的規則すべてを利用して二波形分離を行う際の分離抽出の対象となる音とは、AM–FM 調波複合音であると解釈できる。従って、本論文では二波形分離問題において、分離抽出の対象となる音を AM–FM 調波複合音、もう一つの音を AM–FM 複合音と仮定する。ここで、AM–FM 調波複合音を式 (2.1) に基づき、次式で定義する。

$$\sum_n A_n(t)\exp(j2\pi nF_0t + j\theta_{1n}(t)) \tag{2.3}$$

但し、n は調波成分の次数を示す番号、$A_n(t)$ は瞬時振幅、$\theta_{1n}(t)$ は瞬時位相、F_0は基本周波数とする。本来、基本周波数 F_0は時間関数 $F_0(t)$ で表現されるべきだが、ここでは一定の F_0を中心に周波数変調項が位相変調の形式で表されるものとする。また、n に対する$\sum$は調波関係に基づいた総和計算である。この定義から、本論文では振幅情報、位相情報、基本周波数の三つの物理量を取り扱う。

次に、AM–FM 調波複合音が実際にどのような自然界の音を表現できるか、その網羅性を議論する。

2.3.4 AM–FM 調波複合音の網羅性

図 2.3に音源とそれによって生じる音の構成を示す。我々が一般に観測する音 $s(t)$ とは、音源波 $e(t)$ で生じた音が何らかの観測フィルタを通過した音（例えば共振器を通過した音）

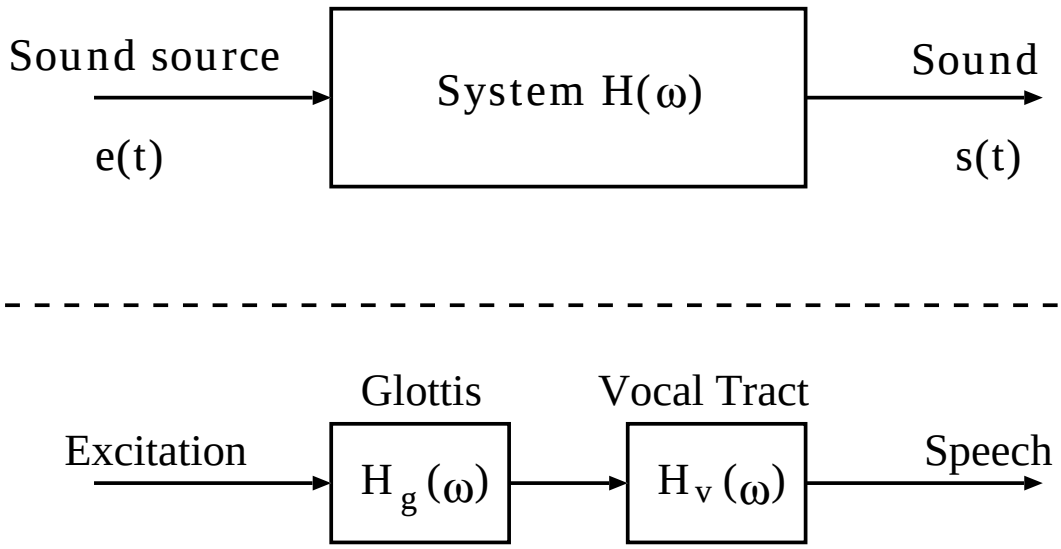

図 2.3: 観測信号の生成過程

である。

そこで、発見的規則 (iii) でも述べられたように、音源波 $e(t)$ を繰り返し周期で振動する音とし、

$$e(t) = \sum_n e_n \exp(j\omega_n t + j\eta_n) \tag{2.4}$$

と仮定する。但し、$\omega_n = nF_0$である。この仮定に基づく音原波は、音声でいうと有声音を駆動するスパイクトレインであり、弦や面（分割モードを生じるものは除く）で振動するモードに相当する。ここで、$e(t)$ の Fourier 変換を $E(\omega)$ とすれば、

$$\begin{aligned} E(\omega) &= \int_{-\infty}^{\infty} \left(\sum_n e_n \exp(j\omega_n t + j\eta_n) \right) \exp(-j\omega t) dt \\ &= \sum_n e_n \int_{-\infty}^{\infty} \exp(j\omega_n t + j\eta_n) \exp(-j\omega t) dt \\ &= 2\pi \sum_n e_n \delta(\omega - \omega_n) \exp(j\eta_n) \end{aligned} \tag{2.5}$$

となる。次に、音源波が、共振器などの環境を表すフィルタを通過して観測音になると考える。ここで、フィルタ特性を

$$H(\omega) = h(\omega) \exp(j\epsilon(\omega)) \tag{2.6}$$

とする。但し、$h(\omega)$ は、$\omega = \omega_n = n \cdot f_0$で共振点をもつような特性、つまり$\omega = \omega_n$で極をもつ帯域通過フィルタ特性である。このとき、観測信号 $s(t)$ の Fourier 変換 $S(\omega)$ は

$$S(\omega) = E(\omega)H(\omega) = 2\pi \sum_n e_n h(\omega_n) \delta(\omega - \omega_n) \exp(j\eta_n + j\epsilon(\omega_n)) \tag{2.7}$$

となる。次に、上式を逆 Fourier 変換すると

$$\begin{aligned} s(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} 2\pi \sum_n e_n h(\omega_n)\delta(\omega-\omega_n)\exp(j\eta_n + j\epsilon(\omega_n) + j\omega t)d\omega \\ &= \sum_n e_n h(\omega_n)\exp(j\omega_n t + j\eta_n + j\epsilon(\omega_n)) \\ &\approx \sum_n A_n(t)\exp(j\omega_n t + j\theta_n(t)) \end{aligned} \tag{2.8}$$

を得る。但し、$A_n(t) = e_n h_n(t)$、$\theta_n(t) = \eta_n + \epsilon_n(t)$ であり、$h_n(t)$ と $\epsilon_n(t)$ はそれぞれ、$h(\omega_n)$ と $\epsilon(\omega_n)$ を時間関数に変換したもの（逆 Fourier 変換）とする。この表現から、本論文で取り扱う AM–FM 調波複合音は音源波が周期的に振動するものであり、それは式 (2.4) で表現できるような音源波をもつものである。例えば、式 (2.4) の音源波を利用し、システム関数 $H(\omega)$ を声帯フィルタ $H_g(\omega)$ と声道フィルタ $H_v(\omega)$ の積とすれば、有声音（母音）の生成メカニズムを解析できる。また、弦や面（分割モードが生じる場合を除く）のように、一般に振動モードが整数になるものは調波複合音になるため、ギターやピアノといった楽器音も上記の形式で表現できる。従って、本論文で取り扱う分離抽出の対象となる音は、音声や楽器音といった音源波が調波構造を生む音すべてを表現できる。

以上、分離抽出の対象となる音の網羅性を議論したが、次に二波形分離問題を解くために利用する制約条件の概念を述べる。

2.3.5 二波形分離問題で利用する制約条件の概念

前節で述べたように、二波形分離問題における分離抽出の対象となる音とは、次のように構成される AM–FM 調波複合音とした。

1. 各高調波は立上りと立下りで同期する。

2a. 高調波を構成する単一成分音は AM 単一成分音（振幅変調された正弦波信号）である。

2b. 基本周波数は時間的に変動し（周波数変調される）、各高調波の周波数もこれに対応して変動する（周波数変調される）。

3. 各高調波は基本周波数の整数倍で構成される。

4. 調波複合音は同様に振幅変調される。

ここで、上記の構成順はそれぞれ発見的規則の順番に対応している。この AM–FM 調波複合音を単一成分音、AM 調波複合音、AM-FM 調波複合音の順に構成を複雑にした場合を考えてみる。AM 単一成分音の場合は、単一成分の周波数以外に音が存在しないため、上

表 2.1: AM–FM 調波複合音の構成とそれに対応する発見的規則

音の構成	発見的規則 (i)	発見的規則 (ii)	発見的規則 (iii)	発見的規則 (iv)
AM 単一成分音	×	○	×	○
AM 調波複合音	○	○	○	○
AM–FM 調波複合音	○	○	○	○

○：有効、×：無効

記の 1. と 3. に該当しない。しかし、振幅変調による変化はあるのだから、上記の 2. と 4. に該当する。また、残る二つの調波複合音の場合は、上記の四つの項目すべてに該当する。従って、AM–FM 調波複合音を構成する三つの音それぞれに対しては、表 2.1に示す発見的規則を利用できることがわかる。しかし、二波形分離問題では、上記に示した AM–FM 調波複合音と、もう一つの AM–FM 複合音（どんな音かわからない）が加算された状態から、AM–FM 調波複合音を分離抽出する問題となるため、利用する制約条件の働きを考えなければならない。

そこで、次に Bregman によって提唱された四つの発見的規則とこれを二波形分離問題の枠組に対応させた場合について考察する。このときの概念を図 2.4に示す。また、この制約条件の概念に拘束される AM–FM 調波複合音の構成を図 2.5に示す。

はじめに、図 2.4(a) に示す発見的規則 (i) の解釈を説明する。もし、複合音を構成する各単一成分音の立上り・立下りが同期していれば、この複合音は一つの音源で生じた音として知覚するし、そうでなければ、異なる複数の音として知覚する。従って、二波形分離問題には二つの音源が存在するのだから、分離抽出したい音の立上り・立下りを拘束することで二波形を分離する。式 (2.3) に基づけば、AM–FM 調波複合音を分離抽出するために、瞬時振幅 $A_n(t)$ と瞬時位相 $\theta_{1n}(t)$ をこの規則で拘束すればよいことがわかる。このことから、発見的規則 (i) は物理量を直接拘束する制約条件というよりは、分離抽出をするための検出に役立つ制約条件であるといえる。

次に、図 2.4(b) に示す発見的規則 (ii) の解釈を説明する。もし、音のパワーが時間的にも周波数的にも連続であれば、それは一つの音源で生じた音として知覚するし、そうでなければ異なる複数の音として知覚する。従って、二波形分離問題では、分離抽出したい音の瞬時振幅や瞬時位相、基本周波数の時間変化を拘束することで二波形を分離する。式 (2.3) に基づけば、AM–FM 調波複合音を分離抽出するために、瞬時振幅 $A_n(t)$ と瞬時位相 $\theta_{1n}(t)$ をこの規則で拘束すればよいことがわかる。ここで、図 2.5から予想できるように、振幅あるいは位相の時間変化は中心角周波数を中心とした帯域幅で拘束されてしまう。つまり、

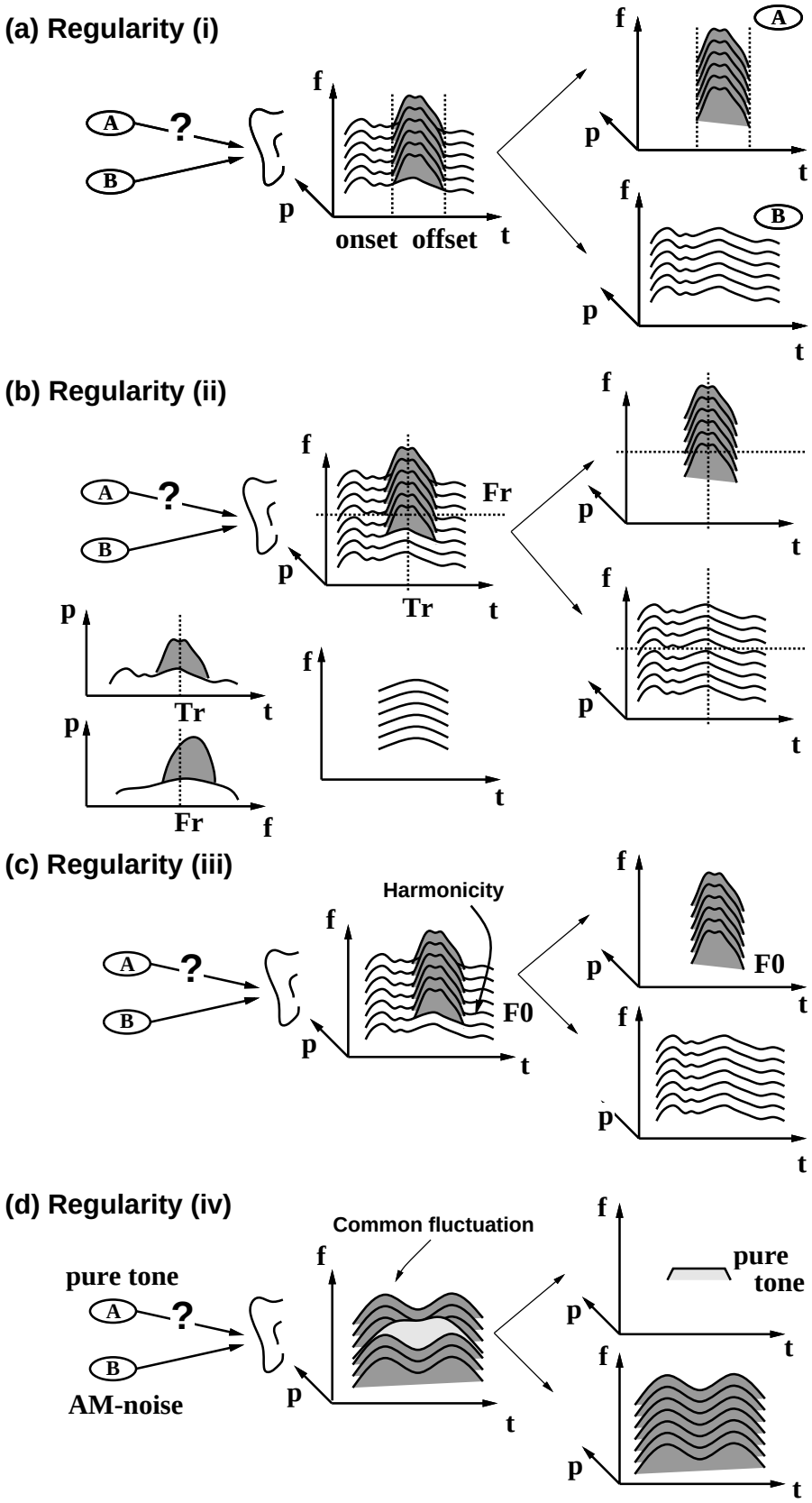

図 2.4: Bregman によって提唱された四つの発見的規則. (a) 共通の立上り・立下りの規則, (b) 漸近的変化の規則, (c) 調波関係の規則, (d) 一つの音響事象に生じる変化に関する規則.

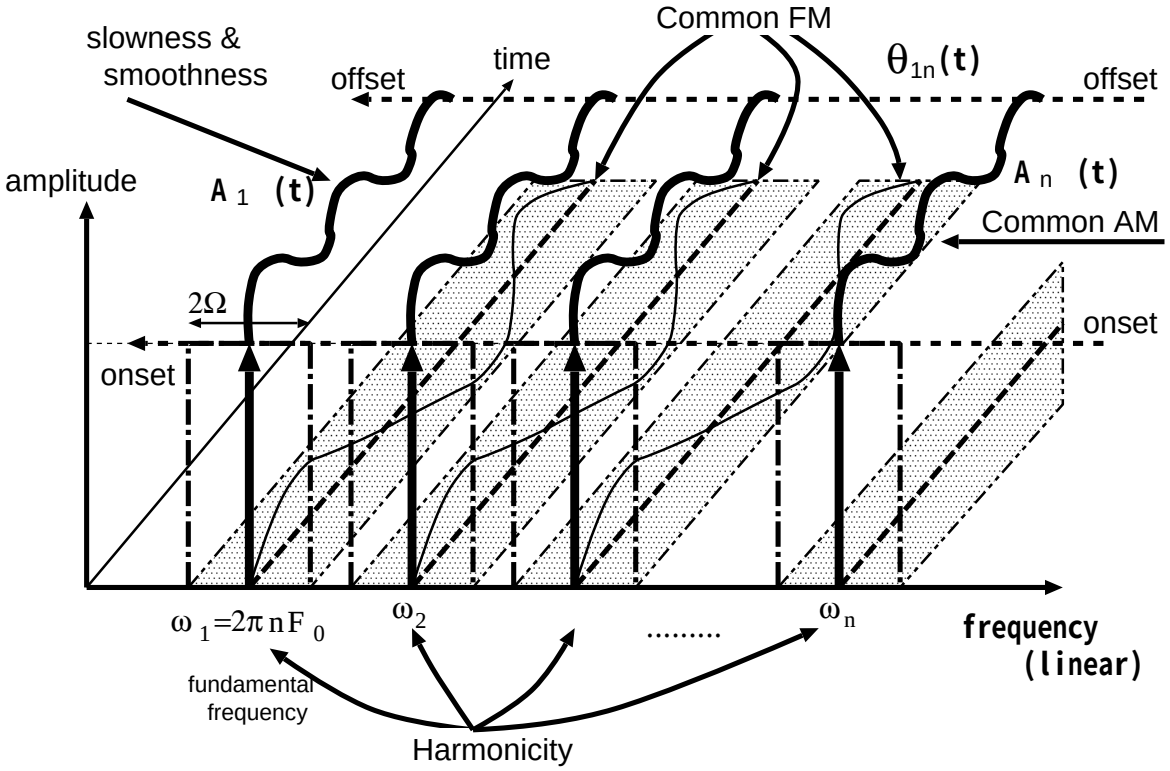

図 2.5: 四つの発見的規則で拘束される AM–FM 調波複合音

連続性（漸近的変化）の制約条件は、振幅と位相の時間変動を許容する最大周波数を拘束することになる。また、式 (2.3) の形式を満たすために、振幅の時間変動を許容する最大周波数は、その中心角周波数に対して極めて小さい値でなければならない。このことから、発見的規則 (ii) は物理量を直接拘束する制約条件であるといえる。

次に、図 2.4(c) に示す発見的規則 (iii) の解釈を説明する。もし、高調波成分が基本周波数の整数倍に構成されているのであれば、一つの音源で生じた音として知覚するし、そうでなければ異なる複数の音として知覚する。従って、二波形分離問題では、分離抽出したい音の調波構造を拘束することで二波形を分離する。式 (2.3) に基づけば、AM–FM 調波複合音を分離抽出するために、瞬時振幅 $A_n(t)$ と瞬時位相 $\theta_{1n}(t)$ に対する中心角周波数を拘束すればよい。このことから、発見的規則 (iii) は物理量を直接拘束する制約条件というよりは、分離抽出をする信号成分の検出に役立つ制約条件であるといえる。また、図 2.5からもわかるように、中心角周波数と連続性によって拘束される帯域幅の値によっては、周波数帯域で隙間ができる場合とできない場合がある。

最後に、図 2.4(d) に示す発見的規則 (iv) の解釈を説明する。もし、音の振幅が同じように変動しているのであれば、一つの音源で生じた音と知覚するし、そうでなければ異なる

複数の音として知覚する。従って、二波形分離問題では、分離抽出したい音の振幅包絡間の共通の変動を拘束することで二波形を分離する。式 (2.3) に基づけば、AM–FM 調波複合音を分離抽出するために、瞬時振幅 $A_n(t)$ をこの規則で拘束すればよいことがわかる。また、位相の時間変化については、振幅包絡を一定であると考えれば、位相の時間変化を振幅の時間変化として取り込むことができる。これは、母音の高調波の振幅包絡に基本周波数成分が振幅変調されている現象を説明する一例である。ここで、図2.5から予想できるように、振幅の時間変化のコヒーレンスは各周波数成分間における振幅包絡の相関値で評価できる。以上のことから、発見的規則 (iv) は物理量を直接拘束する制約条件であるといえる。また、この発見的規則 (iv) は、各中心角周波数における帯域幅を一定にする、つまり振幅と位相の時間変動を各周波数成分間で統一することを意味する。

以上、二波形分離問題で利用する制約条件の概念を述べた。先に示したように、ここで取り扱う音の物理量は、音の振幅情報、位相情報、そして基本周波数（基本波）である。しかし、歴史的な背景をみると、計算論的な聴覚の情景解析の研究や数理工学的な研究といった他の研究では、物理量として振幅（あるいはパワー）情報や基本周波数しか扱っていない。これはおそらく位相情報が取り扱い難いということと、一般に聴覚は位相に対して聾であるといわれてきたことに起因するかもしれない。しかし、Bregman は音響事象が音の振幅（あるいはパワー）情報であるとは断言していない。また、図 2.4 (d) のように、同一周波数軸上に二つの信号成分が加算されているような場合、位相を使わない限り完全に二つに分離することは難しい。例えば、振幅情報のみを利用しているのならば、二つの和を 10 としたとき、$5 + 5 = 10$ という場合もあるし、$2 + 8 = 10$ という場合もある。両者の差を表すものが位相情報であり、二つの信号が混合されたときに失われる情報でもある。この事実を踏まえ、本論文では二波形分離問題に対し、各発見的規則を利用して音を分離する状況を考える。

2.4 発展的構築法

次に、音の分離抽出における聴覚の計算の方略を構築するための方法として発展的構築法を提案する。具体的には、以下の手順に従って、音の分離抽出における聴覚の計算の方略の構築を試みる。

1. 信号音を、振幅変調（AM）および周波数変調（FM）された成分をもつ AM–FM 複合音として物理的に表現する。

2. 二つの AM–FM 複合音の和として二波形分離問題の定式化を行なう。但し、分離抽出の対象となる音は AM–FM 調波複合音であり、もう一つの音は AM–FM 複合音である。

3. 二波形分離問題を一意に解くための制約条件として、Bregman によって提唱された四つの発見的規則を定式化し、二波形分離問題の解法を提案する。

4. 二波形分離問題の解法において、制約条件の十分性を検証する。ここでは、瞬時振幅、瞬時位相、基本周波数に対する分離精度を評価することで制約条件の十分性を検証するため、分離抽出の対象となる音を

   (a) AM 単一成分音：正弦波信号 or AM 単一成分音

   (b) AM 調波複合音：一定な基本周波数をもつ調波複合音（AM のみ）

   (c) AM-FM 調波複合音：合成母音

   の順序で利用する。これにより、二波形分離問題の解法に対し、(a) から瞬時振幅の分離精度、(b) から瞬時位相の分離精度、(c) から基本周波数の時間変動による影響についてその十分性を検証する。

5. AM–FM 調波複合音を用いて二波形分離問題の解法で利用した制約条件の有効性を検証する。ここでは、制約条件を一つずつ省略した場合の分離精度を評価することで、制約条件の有効性を検証する。

6. 以上の検証を踏まえ、二波形分離問題における音の分離抽出の計算の方略を導出する。

以上のように、AM 単一成分音から AM–FM 調波複合音へと発展的に拡張した信号を利用して、二波形分離問題の解法を検証することから、本論文で提案した構築方法を発展的構築法と呼ぶ。

本論文では、最終的に AM-FM 調波複合音の分離抽出が可能となる二波形分離モデルを実現した後で、計算の方略を (1) 実音声（母音）を対象にした二波形分離問題、(2) 共変調マスキング解除を想定した二波形分離問題に展開することで、この計算の方略がこれらの問題の解法を導出できることを実証する。

2.5 むすび

本章では、音の分離抽出における聴覚の計算の方略を構築するための方法論を提案した。

はじめに、「二つの波形を分離する」という基本的な機能に着目し、不良設定問題である信号分離問題を二波形が加算されたものから個々の音に分離抽出する、二波形分離問題とした。

次に、本論文で取り扱う信号音を sinusoidal model で知られる AM–FM 複合音の物理的表現で定義した。その後で、Bregman によって提唱された四つの発見的規則を解釈することで、二波形分離の対象となる、自然界に存在する音を AM–FM 調波複合音と仮定した。これは、主に発見的規則を分離することの規範と見なした時に、本論文で取り扱う音とはどのようなものかを述べたものである。具体的には、AM–FM 調波複合音を以下のように定義した。

1. 各高調波は立上りと立下りで同期する。

2a. 高調波を構成する単一音は AM 単一成分音（振幅変調された正弦波信号）である。

2b. 基本周波数は時間的に変動し（周波数変調されている）、各高調波の周波数もこれに対応して変動する（周波数変調される）。

3. 各高調波は基本周波数の整数倍で構成される。

4. 調波複合音は同様に振幅変調される。

また、AM 単一成分音の場合は、音が単一成分音で構成されることから複合成分間の制約を必要とせず、四つの発見的規則のうち、発見的規則 (i) と (iii) は無効であると考えた。次に、ここで仮定した AM–FM 調波複合音の網羅性を議論し、二波形分離問題を解くために利用する制約条件の概念を述べた。

最後に、音の分離抽出における聴覚の計算の方略を構築するための方法論として、発展的構築法を提案した。この発展的構築法とは、AM–FM 調波複合音を構成する AM 単一成分音から AM 調波複合音、周波数変調された AM 調波複合音（AM-FM）という順序に拡張した信号を利用し、二波形分離問題の解法で利用する制約条件の十分性および有効性を検証することである。このように、分離抽出音を発展的に拡張して検証することから、この方法は発展的構築法と名付けられた。

第 3 章

二波形分離問題の理論的検討

3.1 まえがき

本章では、発展的構築法に従い、音の分離抽出における聴覚の計算理論を構築するための最初の手続きとして、音の振幅情報と位相情報を利用する二波形分離問題の解法を提案する。

はじめに、前章で定義した AM–FM 調波複合音を分離対象として取り扱える二波形分離問題を定式化する。ここで扱う音の物理量は、瞬時振幅、瞬時位相、基本周波数の三つである。

次に、定式化した二波形分離問題が不良設定問題であることを理論的に示す。これは、観測された混合信号の瞬時振幅と瞬時位相から、二波形の瞬時振幅と瞬時位相を同時にかつ一意に決定できないことを示すものである。このとき、不良設定問題を一意に解くために利用する制約条件とその定式化も述べる。これは、前章で示した四つの発見的規則をどのように表記するかということとそれに対する音の構成に深く関係するものである。

次に前章で述べた二波形分離における物理量をどのように取り扱うか述べる。特に、聴覚末梢系における聴覚フィルタ群の特性と、音の物理量として振幅と位相、基本周波数の取り扱いについて説明する。

次に、二波形分離問題を解くためのアルゴリズムと信号処理の流れを説明する。特に、各ブロックにおける処理例として、分析合成フィルタ群の変換・逆変換の結果（フィルタバンクを素通りさせた結果）と、基本周波数推定部のロバスト性を示す。

最後に、本章で提案した二波形分離問題の解法が混合信号から AM–FM 調波複合音を分離抽出できることを示す。

3.2 二波形分離問題の定式化

本論文では、前章で述べた sinusoidal model に基づき、取り扱う信号音を AM–FM 複合音で表現する。そのため、観測した混合信号の分離問題を AM–FM 複合音の和で表現される問題として定義する。

はじめに、ある二つの音響信号 $f_1(t)$ と $f_2(t)$ が $f(t) = f_1(t) + f_2(t)$ に加算され、混合信号 $f(t)$ のみを受音できるものとする。ここで、$f_1(t)$ を望みの信号、$f_2(t)$ を雑音あるいはそれ以外の音とする。これは、図 3.2に示す K個の分析（聴覚）フィルタ群により周波数分解される。ここで、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

$$X_{1,k}(t) \quad = \quad A_k(t) \exp(j\omega_k t + j\theta_{1k}(t)) \tag{3.1}$$

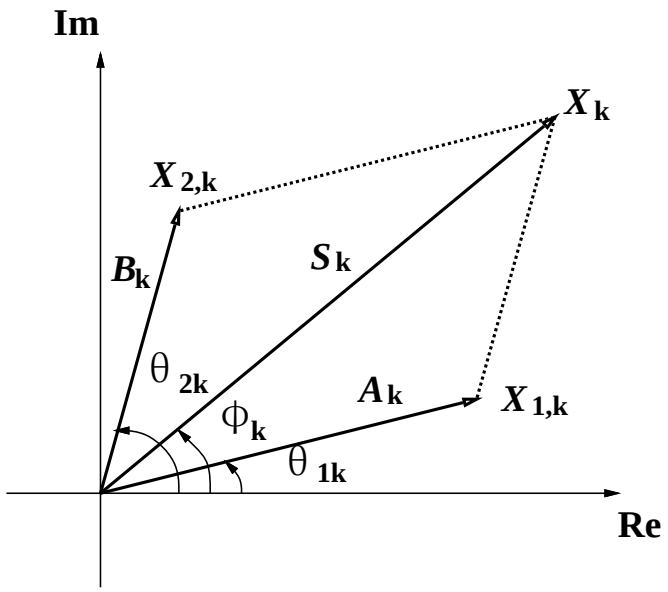

図 3.1: 複素スペクトルのベクトル図

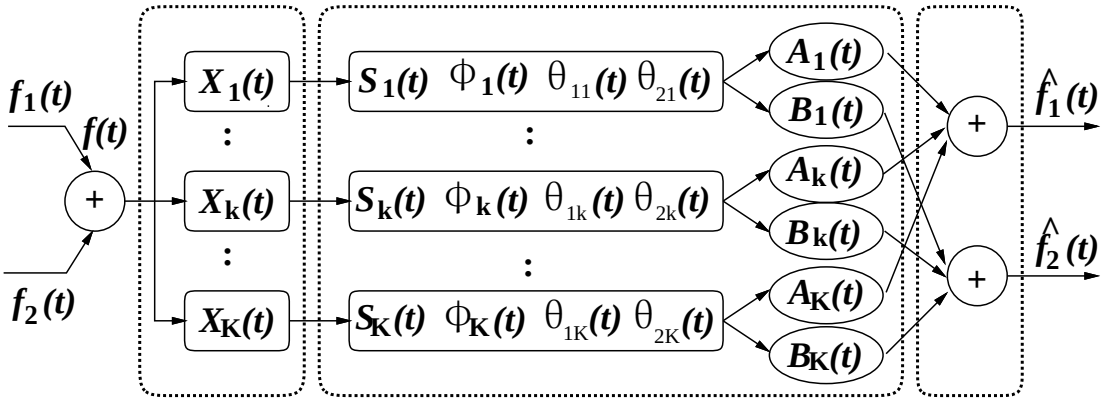

図 3.2: 二波形分離問題の定式化

$$X_{2,k}(t) = B_k(t)\exp(j\omega_k t + j\theta_{2k}(t)) \tag{3.2}$$

と仮定すれば、$f(t)$ の通過成分 $X_k(t)$ は、

$$X_k(t) = X_{1,k}(t) + X_{2,k}(t) \tag{3.3}$$

$$= S_k(t)\exp(j\omega_k t + j\phi_k(t)) \tag{3.4}$$

と表される。但し、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、時刻t における $X_k(t)$ に着目すれば、$X_{1,k}(t)$ と $X_{2,k}(t)$ との関係は、図 3.1で表される。また、$S_k(t)$ と $\phi_k(t)$ は、それぞれ、式 (3.1)〜(3.4) から

$$S_k(t) = \sqrt{A_k^2(t) + 2A_k(t)B_k(t)\cos\theta_k(t) + B_k^2(t)} \tag{3.5}$$

$$\phi_k(t) = \arctan\left(\frac{A_k(t)\sin\theta_{1k}(t) + B_k(t)\sin\theta_{2k}(t)}{A_k(t)\cos\theta_{1k}(t) + B_k(t)\cos\theta_{2k}(t)}\right) \tag{3.6}$$

で求められるため、$A_k(t)$ と $B_k(t)$ は、それぞれ

$$A_k(t) = \frac{S_k(t)\sin(\theta_{2k}(t) - \phi_k(t))}{\sin\theta_k(t)} \tag{3.7}$$

$$B_k(t) = \frac{S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))}{\sin\theta_k(t)} \tag{3.8}$$

として解くことができる。但し、$\theta_k(t) = \theta_{2k}(t) - \theta_{1k}(t)$ であり、$\theta_k(t) \neq n\pi, n \in \mathbf{Z}$ とする。同様に上式を整理すると、$\theta_{1k}(t)$ と$\theta_{2k}(t)$ は、それぞれ

$$\theta_{1k}(t) = \arcsin\left(\frac{A_k(t)Y_k(t)}{S_k(t)\sqrt{Y_k(t)^2+1}}\right) - \arctan\left(\frac{Y_k(t)\cos\phi_k(t) - \sin\phi_k(t)}{Y_k(t)\sin\phi_k(t) + \cos\phi_k(t)}\right) \tag{3.9}$$

$$\theta_{2k}(t) = \arcsin\left(-\frac{B_k(t)Y_k(t)}{S_k(t)\sqrt{Y_k(t)^2+1}}\right) - \arctan\left(\frac{Y_k(t)\cos\phi_k(t) + \sin\phi_k(t)}{Y_k(t)\sin\phi_k(t) - \cos\phi_k(t)}\right) \tag{3.10}$$

として解くことができる（導出過程は付録 A 参照)。但し、

$$Y_k(t) = \frac{\sqrt{(2A_k(t)B_k(t))^2 - Z_k(t)^2}}{Z_k(t)} \tag{3.11}$$

$$Z_k(t) = S_k(t)^2 - A_k(t)^2 - B_k(t)^2 \tag{3.12}$$

である。しかし、上記の定式化において、観測された混合信号の瞬時振幅 $S_k(t)$ と瞬時出力位相 $\phi_k(t)$ 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。本論文では、Bregman によって提唱された四つの発見的規則を制約条件として用

いて、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考える。

最後に、すべての分析フィルタの出力において、二波形の瞬時振幅と瞬時入力位相を求めた後、分析フィルタ群と逆の操作を行うことで、$f_1(t)$ と $f_2(t)$ をそれぞれ再構成する。ここで、再構成された信号をそれぞれ$\hat{f}_1(t),\ \hat{f}_2(t)$ とする。

3.3 二波形分離問題の解法

3.3.1 二波形分離問題における仮定

本論文では、$f_1(t)$ を分離抽出の対象となる AM–FM 調波複合音、$f_2(t)$ を妨害雑音とし、$f_2(t)$ 中に $f_1(t)$ が加算される状態から、$f_1(t)$ を分離抽出する二波形分離問題とする。また、この調波複合音は、基本周波数 $F_0(t)$ を整数倍した調波関係を満たす高調波成分を持つものとする。

さて、本章で定義された二波形分離問題は二つの sinusoidal model の和として定義されている。従って、sinusoidal model における瞬時振幅と瞬時位相の推定方法（例えば、ピークピッキングやハーモニックトラッキング）を二波形分離問題の解法として応用できる。しかし、この推定方法では、一つの音ないし、周波数領域で独立な二つの音に対して有効であるが、二つの信号成分が同一周波数領域で重複するような場合、厳密な分離をすることが難しい。そのため、一意な解を求めるにあたり、sinusoidal model で利用されてきた方法を利用することはここでは好ましくない。そこで、本論文では不良設定問題である二波形分離問題を最適化の規範で解くために、分離抽出したい信号の瞬時振幅、瞬時位相、および基本周波数の時間変化に注目して望みの信号を分離抽出しているものとする。また、この仮定に基づき、Bregman によって提唱された四つの発見的規則に対応する制約条件を定式化する。

3.3.2 二波形分離問題で利用する制約条件

前章で述べたように、Bregman は聴覚が利用している四つの発見的規則を提唱した。この規則は、我々の経験する音環境に存在する音とはどういうものなのかを述べているのに等しい。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することができない。そこで、四つの発見的規則を表 3.1に示すような関係で制約条件を対応づけ、以下のように定式化する。

表 3.1: Bregman 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
(i) 関連の無い音が一緒に始まったり、終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	漸近的変化
(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音の属性は、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的成分が発生する	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

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 $T_{\rm S}$、立下り時刻を $T_{\rm E}$ とする。このとき、同じ音源で生じた信号成分であれば、高調波成分の立上り $T_{k,\rm on}$ と立下り $T_{k,\rm off}$ 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

$$|T_{\rm S} - T_{k,\rm on}| \leq \Delta T_{\rm S} \tag{3.13}$$

$$|T_{\rm E} - T_{k,\rm off}| \leq \Delta T_{\rm E} \tag{3.14}$$

を満たさなければならない。

■

制約条件 2 (漸近的変化（多項式近似）) ある区間における瞬時振幅 $A_k(t)$、瞬時入力位相 $\theta_{1k}(t)$、基本周波数 $F_0(t)$ のそれぞれの導関数が、

$$dA_k(t)/dt = C_{k,R}(t) \tag{3.15}$$

$$d\theta_{1k}(t)/dt = D_{k,R}(t) \tag{3.16}$$

$$dF_0(t)/dt = E_{0,R}(t) \tag{3.17}$$

で表されるものとする。但し、$C_{k,R}(t),\ D_{k,R}(t),\ E_{0,R}(t)$ は、区分的に微分可能な R次多項式である。このとき、$A_k(t),\ \theta_{1k}(t),\ F_0(t)$ は、それぞれ、$A_k(t) = \int C_{k,R}(t)dt + C_{k,0}$, $\theta_{1k}(t) = \int D_{k,R}(t)dt + D_{k,0}$, $F_0(t) = \int E_{0,R}(t)dt + E_{0,0}$ と表される。 ■

制約条件 3 (漸近的変化（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

$$\sigma_A = \int_{t_a}^{t_b} [A_k^{(R+1)}(t)]^2 dt \tag{3.18}$$

$$\sigma_\theta = \int_{t_a}^{t_b} [\theta_{1k}^{(R+1)}(t)]^2 dt \tag{3.19}$$

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 (3.18) の $C_{k,R}(t)$ と式 (3.19) 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

$$n \times F_0(t), \qquad n = 1, 2, \cdots, N_{F_0} \tag{3.20}$$

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 $A_\ell(t)$ に強い相関がなければならない：

$$\frac{A_k(t)}{\|A_k(t)\|} \approx \frac{A_\ell(t)}{\|A_\ell(t)\|}, \qquad k \neq \ell. \tag{3.21}$$

但し、$\|\cdot\|$ はノルム記号である。

■

ここで、制約条件式 (3.15) を式 (3.7) に適用することで、一階線形微分方程式が得られる。これを解くことにより、入力位相差 $\theta_k(t)$ の一般解は、補題 1 より求めることができる。

補題 1 入力位相差 $\theta_k(t)$ の一般解は、

$$\theta_k(t) = \arctan\left(\frac{S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))}{S_k(t)\cos(\phi_k(t) - \theta_{1k}(t)) - C_k(t)}\right) \tag{3.22}$$

で得ることができる。但し、$C_k(t) = \int C_{k,R}(t)dt + C_{k,0}$である。

（証明）付録 B 参照。

■

3.3.3 二波形分離モデルの構成

本論文では、二波形分離問題の解法を計算機上で利用するため、図3.3に示すような二波形分離モデルを実装する。これは、主に、(a) 分析フィルタ群、(b) 基本周波数推定部、(c) 波形分離部、(d) グルーピング部の四つのブロックで構成される。また、二波形分離モデルの大まかな処理の流れは、次のようになる。

1. 混合された信号のみが観測される。
2. 混合信号は周波数分解（瞬時振幅と瞬時位相の成分で表現）される（分析フィルタ群）。
3. 分離抽出する信号の基本周波数を求める（基本周波数推定部）。
4. 分解された振幅と位相から、混合される以前の二波形の瞬時振幅と瞬時位相を求める（波形分離部）。
5. 分離された各振幅と位相から元の二波形に再構成する（グルーピング部）。

3.3.4 二波形分離アルゴリズムの概要

二波形分離アルゴリズムによる信号処理の流れを図3.4に示す。

はじめに、混合信号 $f(t)$ のみが観測され（図3.4. A）、分析フィルタ群により、瞬時振幅 $S_k(t)$ と瞬時位相 $\phi_k(t)$ に分解される（図3.4. B、C）。次に、$S_k(t)$ から基本周波数 $F_0(t)$ を求め（図3.4. D）、二波形分離の対象となる時間–周波数領域を決定する。調波成分の存在する周波数領域については、$F_0(t)$ と発見的規則 (iii) の調波関係（図3.4. E、a–a’ ）を用いて決定する。調波成分の存在する時間領域については、発見的規則 (i) の各高調波成分の立上りと立下りの同期（図3.4. F、b–b’）を用いて決定する。

次に、波形分離部では、上記で決定された時間–周波数領域において $S_k(t)$ と $\phi_k(t)$ から四つのパラメータ（$A_k(t)$, $B_k(t)$, $\theta_{1k}(t)$, $\theta_{2k}(t)$）を求める（図3.4）。これは、$A_k(t)$ と $\theta_{1k}(t)$ を発見的規則 (ii) の漸近的変化（ゆっくりと）を用いて最適化問題として解く（図3.4. H、I）。但し、最適解の候補が多過ぎるため、発見的規則 (ii) の漸近的変化（なめらかさ）を加え、解の探索範囲を狭め、発見的規則 (iv) の振幅包絡の変動の一致（相関）を手がかりとして最適解の絞り込みを行う。

最後に、グルーピング部では、$A_k(t)$ と $\theta_{1k}(t)$ および、$B_k(t)$ と$\theta_{2k}(t)$ がそれぞれグルーピングされ、合成フィルタ群を用いて$\hat{f}_1(t)$ と$\hat{f}_2(t)$ に再構成される（図3.4. J）

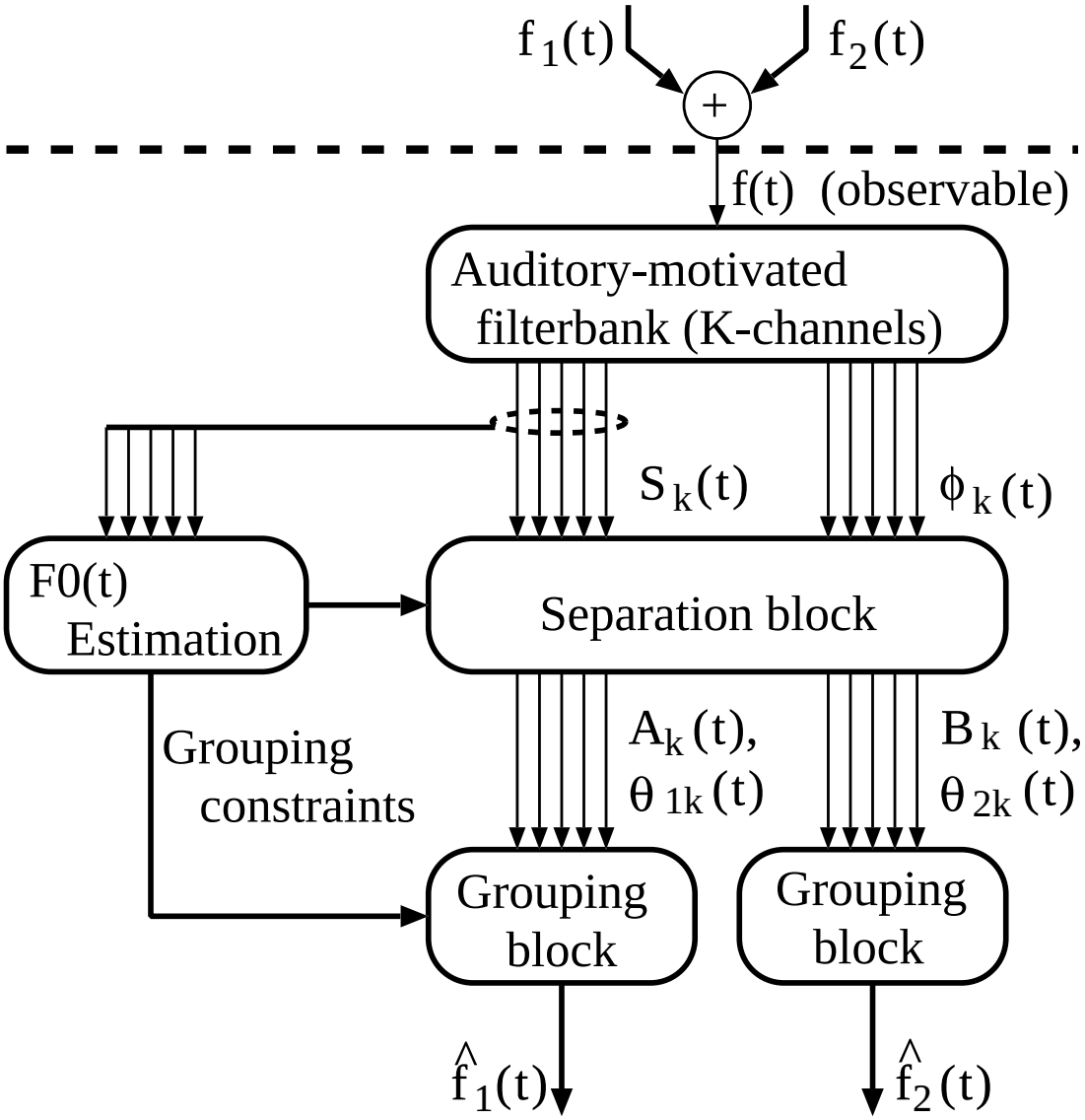

図 3.3: 二波形分離モデル：(a) 分析フィルタ群, (b) 基本周波数推定部, (c) 波形分離部, (d) グルーピング部の四ブロックで構成される

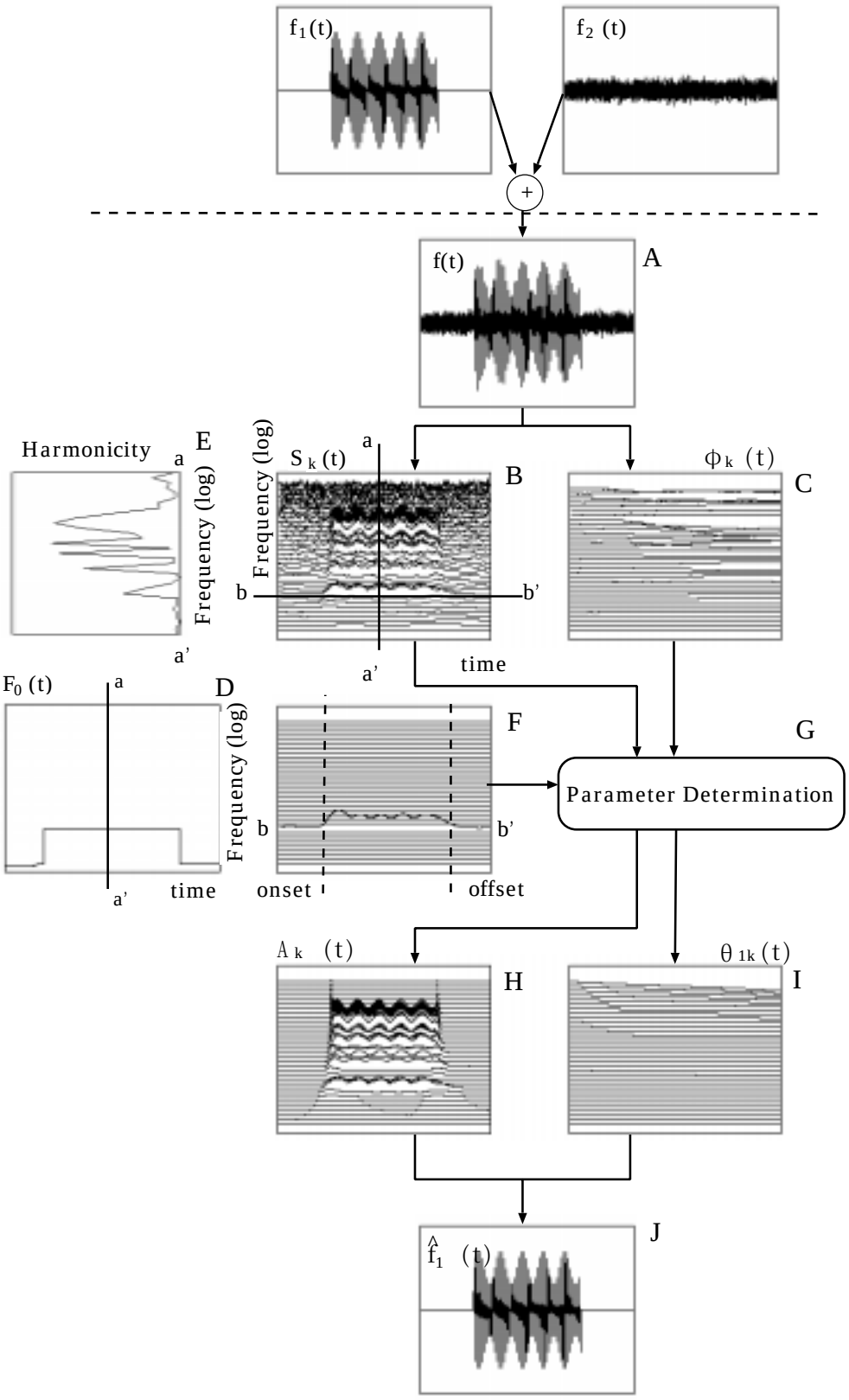

図 3.4: 二波形分離アルゴリズムの概要

3.4 二波形分離アルゴリズムの実装

二波形分離アルゴリズムは図 3.5に示される手順で実行される。はじめに、Step. 1 は、3.4.1 節で説明する定 Q gammatone filterbank（分析フィルタ群）を用いて混合信号を周波数分解する。次に、Step. 2 は、3.4.2 節で説明する計算を行い、Step. 3 は、3.4.3 節で説明する Comb filtering の方法で基本周波数を推定する。次に、Step. 4〜6 および Step. 10 は、3.4.4 節で説明されるグルーピング部を実行する。Step. 7〜9 は、3.4.5 節で説明される波形分離部を実行する。各ステップの詳細の処理は、次節で詳細に述べられる。

3.4.1 分析フィルタ群の実装

本節では、wavelet 変換対（詳細は付録 C 参照）を利用して分析合成系を構築する。本論文では、

1. 聴覚特性を考慮できる分析合成系であること
2. 複素スペクトルを扱えて、かつ不連続点の検出が容易であること

を考慮して、聴覚フィルタのインパルス応答をアナライジング wavelet とした wavelet 分析合成系（定 Q フィルタバンク）を利用する。

聴覚フィルタの特性

最近、聴覚フィルタモデルとして、線形の gammatone filter [Patterson *et al.*, 1995] が利用されている。この聴覚フィルタは、Patterson によって設計された聴覚フィルタであり、基底膜の特性をより良く模擬したものとして知られている。このインパルス応答は、

$$g_t(t) = a_f t^{N_f-1} \exp(-2\pi b_f \mathrm{ERB}(f_c)t)\cos(2\pi f_c t + \varphi), \quad t \geq 0 \tag{3.23}$$

である。但し、a_f, b_f, N_fはパラメータであり、$a_f t^{N_f-1} \exp -2\pi b_f \mathrm{ERB}(f_c)t$ はガンマ分布に関係した振幅項、f_cは搬送波の中心周波数、$\mathrm{ERB}(f_c)$ は中心周波数 f_cにおける等価矩形帯域幅（Equivalent Rectangular Bandwidth）、φ は初期位相である。また、ERB の変換式は

$$\mathrm{ERB}(f) = 24.7\left(\frac{4.37f}{1000}+1\right) \tag{3.24}$$

である [Glasberg and Moore, 1990]。また、この周波数特性は、$f_0 \gg b_f$のとき、近似的に

$$G_T(f) \approx \left[1+\frac{j(f-f_0)}{b_f}\right]^{-N}, \quad 0 < f < \infty \tag{3.25}$$

Step. 1 分析フィルタ群により、$f(t)$ を周波数分解する。

Step. 2 式 (3.27) から瞬時振幅 $S_k(t)$ と式 (3.28) 瞬時出力位相 $\phi_k(t)$ を求める。

Step. 3 瞬時振幅 $S_k(t)$ から基本周波数 $F_0(t)$ を推定する。

Step. 4 基本波成分から立上り $T_{\rm S}$ と立下り $T_{\rm E}$ を求める。

Step. 5 各分析フィルタ出力において立上り $T_{k,\rm on}$ と立下り $T_{k,\rm off}$ を求め、立上りと立下りの同期性を満たす $X_k(t)$ を求める。

Step. 6 基本周波数 $F_0(t)$ の調波関係を満たす $X_k(t)$ を求める。

Step. 7 Step. 5 と Step. 6 のいずれかを満たす時間–周波数領域において、以下の処理を繰り返す。

- **(a)** Kalman filter を用いて、式 (3.15) の $C_{k,0}(t)$ と式 (3.16) の $D_{k,0}(t)$ を推定する。
- **(b)** 推定誤差内 $\hat{D}_{k,0}(t) - Q_k(t) \leq D_{k,1}(t) \leq \hat{D}_{k,0}(t) + Q_k(t)$ から、Spline 補間された $D_{k,1}(t)$ の候補を求める。
- **(c)** 推定誤差内 $\hat{C}_{k,0}(t) - P_k(t) \leq C_{k,1}(t) \leq \hat{C}_{k,0}(t) + P_k(t)$ から、Spline 補間された $C_{k,1}(t)$ の候補を求める。
- **(d)** 式 (3.44) の相関値最大を尺度に、$\hat{C}_{k,1}(t)$ を求める。
- **(e)** (c),(d) を繰り返し、式 (3.46) の相関値最大を尺度に、$\hat{D}_{k,1}(t)$ を決定する。
- **(f)** $\hat{C}_{k,1}(t)$ から$\theta_k(t)$ を、$\hat{D}_{k,1}(t)$ から $\theta_{1k}(t)$ を決定する。これより、$\theta_{2k}(t) = \theta_k(t) + \theta_{1k}(t)$ を決定する。

Step. 9 $S_k(t)$ と$\phi_k(t)$ および、$\theta_{2k}(t)$ と$\theta_{1k}(t)$ から $A_k(t)$ と $B_k(t)$ を求める

Step. 10 グルーピング部により、$A_k(t)$ と$\theta_{1k}(t)$、$B_k(t)$ と$\theta_{2k}(t)$ をそれぞれグルーピングし、$\hat{f}_1(t)$ と$\hat{f}_2(t)$ を再構成する。

図 3.5: 二波形分離アルゴリズム

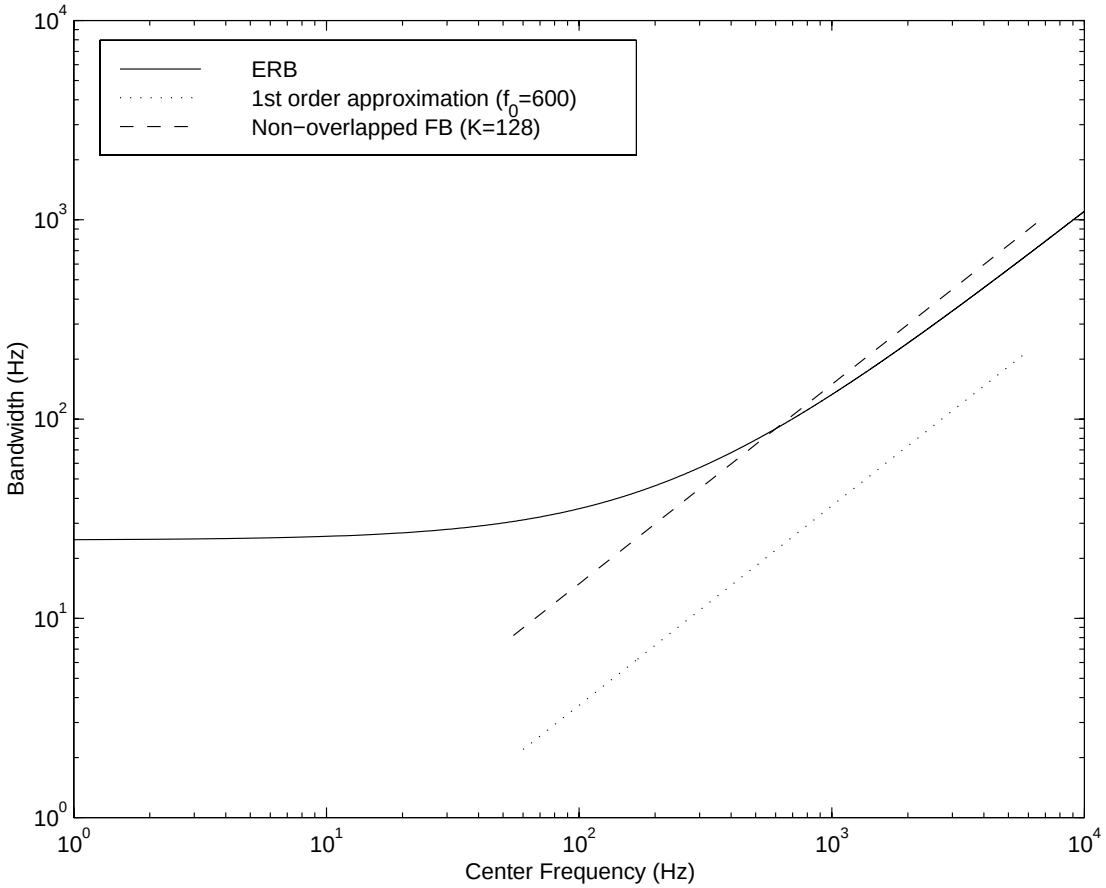

図 3.6: 中心周波数と ERB の関係

と表せる。$G_T(f)$ は、$g_t(t)$ の Fourier 変換を周波数 f の関数で表したものであり、中心周波数を f_0 とする帯域通過フィルタの形態を示している [Patterson and Holdsworth, 1991a]。

wavelet 分析合成系

図 3.6を見ると、おおよそ 800 Hz より高域で ERB を一次式で近似表現できるため、線形フィルタを基底関数とした定 Q フィルタバンクを構築できることが予想できる。

そこで、本論文では、gammatone filter を聴覚フィルタモデルとして採用し、これを基底関数とする分析合成系を構築する。また、位相情報を決定するために、式 (3.23) のインパルス応答の実部と虚部が Hilbert 変換で結ばれるような関数として、アナライジング wavelet を定義する。

$$\psi(t) = a_f t^{N_f - 1} \exp(j2\pi f_0 t - 2\pi b_f \mathrm{ERB}(f_0)t) \tag{3.26}$$

但し、中心周波数 $f_0 = 600$ Hz、$N_f = 4$、$b_f = 0.25$ とした。

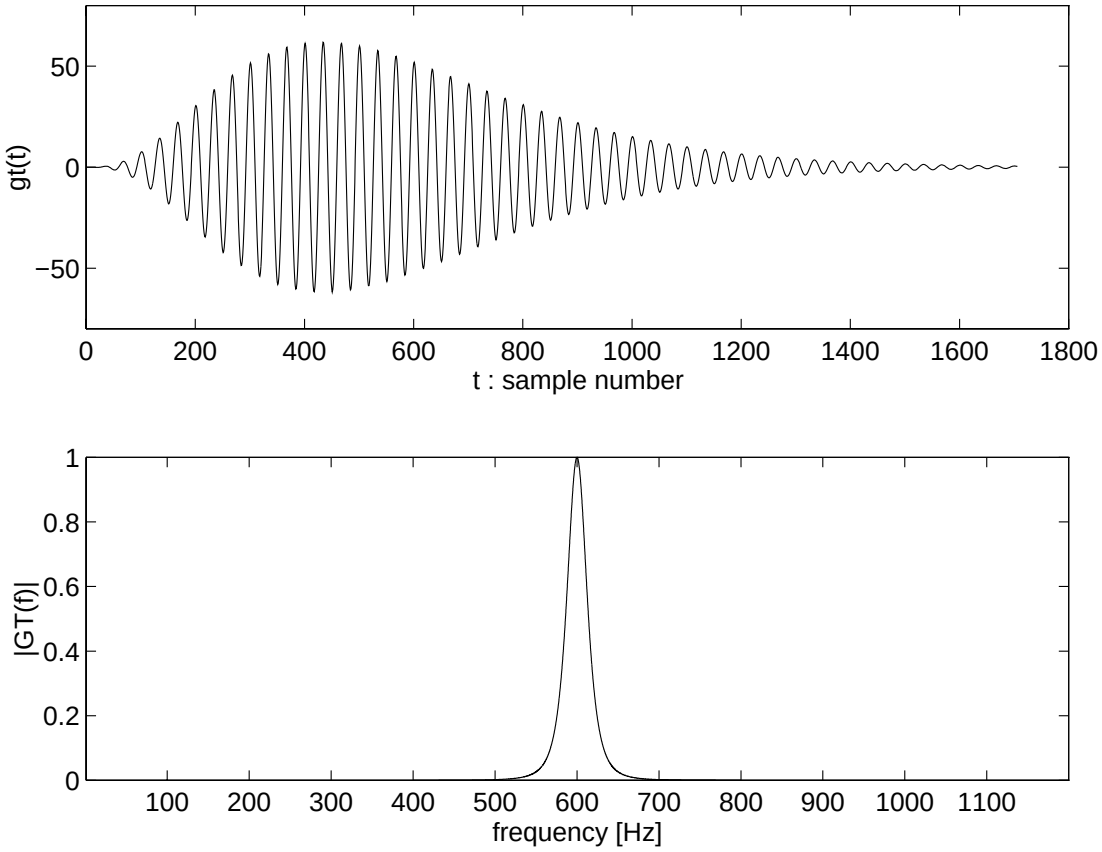

図 3.7: 基本 wavelet $\psi(t)$ の特性：（上）$\mathrm{Re}\{\psi(t)\}$, （下）$\hat{\psi}(f)$（中心周波数 $f_0 = 600$ Hz、$N_f = 4$、$b_f = 0.25$ の gammatone filter）

図 3.7にアナライジング wavelet の時間領域における $\psi(t)$ の実部および周波数領域における$\hat{\psi}(f)$ の振幅特性を示す。この図からもわかるように $GT(0) \approx 0$ となっていることから、式 (3.26) の $\psi(t)$ は許容条件を近似的に満たすことができるので、基本 wavelet として十分利用できることがわかる。

次に、式 (7.20) と式 (7.21) の離散 wavelet 変換を利用し、表 3.2に示す設計仕様で分析合成系を実装した。図 3.8に wavelet 分析合成系の周波数特性を示す。ここで、各フィルタの矩形帯域幅は重複せず、図 3.8のように完全に通過帯域を被覆している。図 3.8（上）は、図 3.6（点線）で設計された分析合成系の特性を示す。また、これは、分析フィルタの矩形幅で定義される帯域幅が重複しないように設計されているため、帯域幅は $K = 128$ で約 1/4 ERB となっている。図 3.8（下）は、$f_0 = 600$ Hz を中心に図 3.6（実線）を一次近似した場合（図中の破線）の特性である。この結果は、$K = 32$、$b_f = 1.019$ で設計した分析合成系法とおおよそ等価である。

工学的な応用を考えた場合、分析フィルタの帯域幅は 1 ERB であることよりも、より

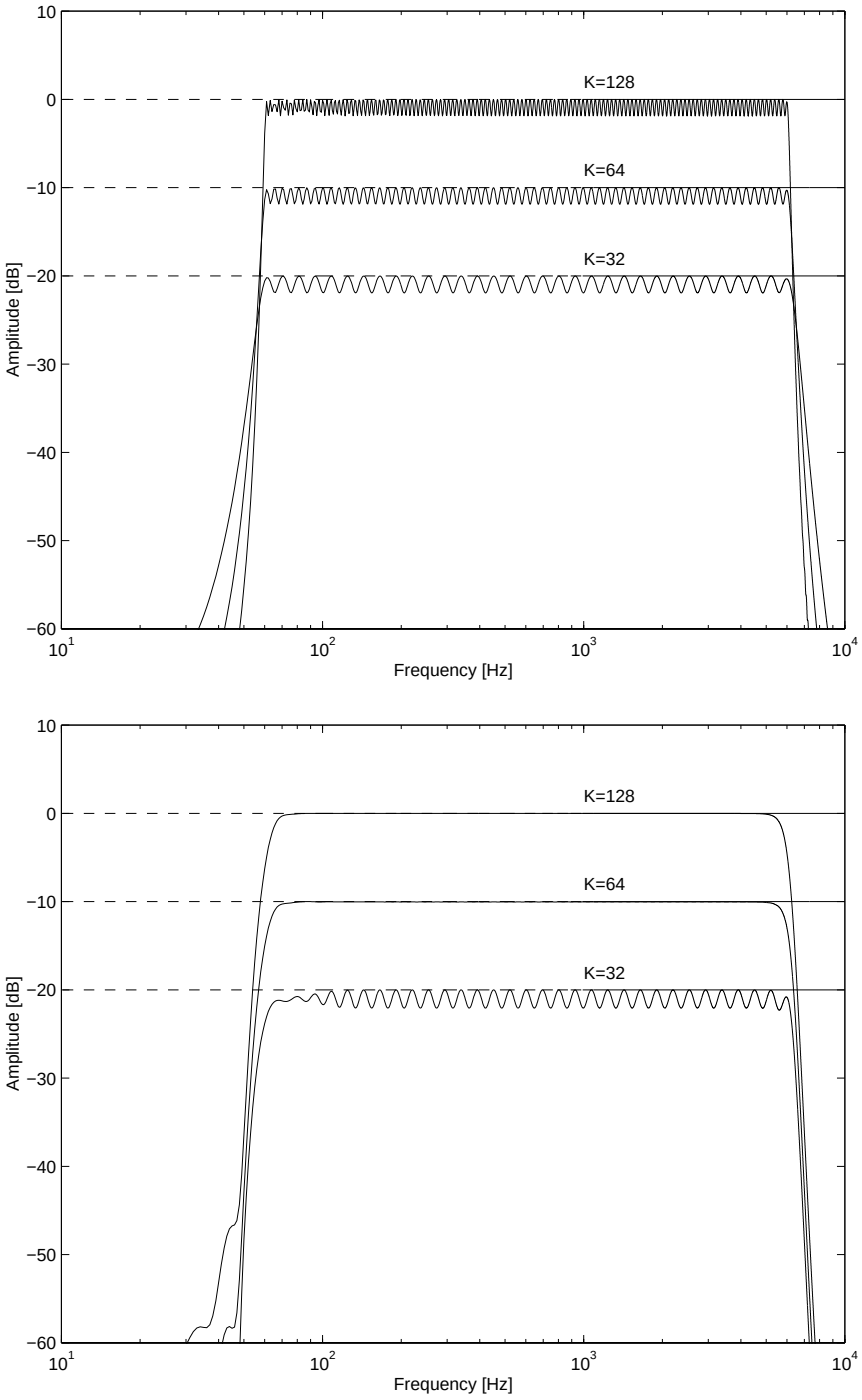

図 3.8: wavelet 分析合成系の周波数特性. 相対レベル：$K = 128$ で 0 dB, $K = 64$ で-10 dB, $K = 32$ で-20 dB）,（上）$b_f = 0.25$ で設計された分析合成系の特性、（下）$b_f = 1.019$ で設計された分析合成系の特性

表 3.2: 分析合成系の設計仕様

記号	定義	設計仕様
f_s	サンプリング周波数	20 kHz
K	フィルタ数	128
W	解析周波数範囲	60〜6000 Hz
a	スケールパラメータ	α^p
α	スケール	$10^{2/K}$
p	インデックス	$-\frac{K}{2} \leq p \leq \frac{K}{2}, p \in \mathbf{Z}$
b_f	帯域幅	$b_f = 0.25$
b	シフトパラメータ	q/f_s
q	インデックス	$q \in \mathbf{Z}$

狭い周波数帯域を分割できるほうが望ましい。そこで、本論文では特に断わりがない限り、$K = 128$、$b_f = 0.25$ で設計された分析フィルタ群を採用することにする。

次に、ここで構築した分析合成系を利用した、音の分解と再構成の一例を図 3.9に示す。図 3.9 (a) は、ATR データベースデータセットにある話者 mau の単母音/a/である [Takeda *et al.*, 1988]。この原信号を $K = 128$ の分析フィルタ群で解析した結果が、図 3.9 (b) である。これは瞬時振幅 $S_k(t)$ の大きさをグレイスケールで表現したものである。但し、この図は各アナライジング wavelet で生じる群遅延を補償するために alignment 処理が施されている。この処理は、式 (3.26) の基底関数の 1 次微分を取り、各スケール軸におけるピークを時間軸（シフト軸）で直線化（alignment）することである。次に、分析フィルタ群の逆の操作を行う合成フィルタ群により、図 3.9 (c) のように再構成される。この再構成においても逆 alignment 処理（上記の説明の逆の操作）が施されている。このとき、原信号と再構成された信号の SNR を測ったところ、約 28 dB であった。この結果から、本分析合成系を利用した信号の分解・再構成については周波数解析範囲内のものであればほぼ完全に復元できることがわかる。

3.4.2 瞬時振幅 $S_k(t)$ と瞬時出力位相 $\phi_k(t)$ の計算方法

次に、wavelet 分析合成系から、瞬時振幅 $S_k(t)$ と瞬時出力位相 $\phi_k(t)$ を求める方法を述べる。式 (3.5) の瞬時振幅 $S_k(t)$ と式 (3.6) の瞬時出力位相 $\phi_k(t)$ は、それぞれ、次の補題 2と 3で得られる。

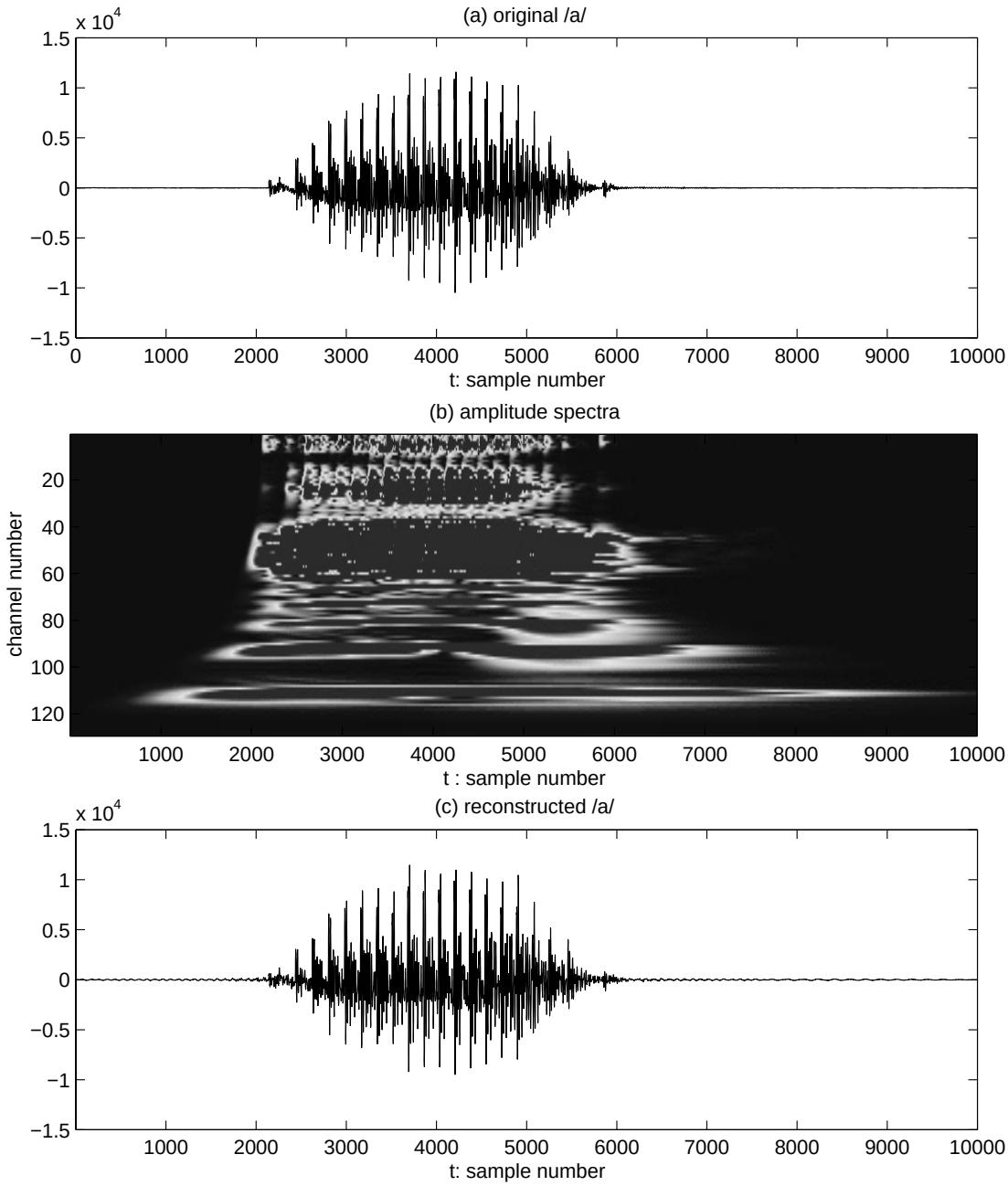

図 3.9: wavelet 分析合成系の評価例：(a) 原音声 /a/, (b) 振幅スペクトログラム（瞬時振幅 $S_k(t)$ のグレイスケール表示）, (c) 再構成された音声信号 /a/

補題 2 瞬時振幅 $S_k(t)$ は、wavelet 変換の振幅項 $|\tilde{f}(a,b)|$ から次式で求めることができる。

$$S_k(t) = |\tilde{f}(\alpha^{k-\frac{K}{2}}, t)|, \qquad a = \alpha^{k-\frac{K}{2}}, b = t \tag{3.27}$$

（証明）付録 D 参照。

■

補題 3 瞬時出力位相 $\phi_k(t)$ は、wavelet 変換の位相項 $\arg(\tilde{f}(a,b))$ から次式で求めることができる。

$$\phi_k(t) = \int_0^t \left(\frac{d}{d\tau} \arg\left(\tilde{f}(\alpha^{k-\frac{K}{2}}, \tau)\right) - \omega_k \right) d\tau, a = \alpha^{k-\frac{K}{2}}, b = t, \tag{3.28}$$

（証明）付録 E 参照。

■

3.4.3 基本周波数の推定方法

本論文では、基本周波数 $F_0(t)$ の推定方法として、比較的雑音にロバストで、分析フィルタ群で推定可能な周波数軸上における Comb filtering を採用する。

はじめに、次のような Comb filter を定義する。

$$\mathrm{Comb}(k,\ell) = \begin{cases} (a+1)/(2f_0 a^{K-k}(a-1)), & \omega_k = n \cdot \omega_\ell \quad 1 \le n \le N \\ 0, & \text{otherwise} \end{cases} \tag{3.29}$$

但し、k, ℓはチャネル番号、Kはチャネル数、Nは最大高調波次数である。次に、時刻 t における Comb filter の通過量を求める。その後で、ℓの探索範囲の上限 L_Fをパラメータとして、通過量を最大とする $\hat{\ell}$を求める。

$$\hat{\ell}(t; L_F) = \arg\max_{\ell \le L_F} \sum_{k=1}^{K} \mathrm{Comb}(k,\ell) S_k(t) \tag{3.30}$$

但し、L_Fは ℓの探索範囲の上限である。次の規範で標準偏差が最小の L_Fにより得られた、$\hat{\ell}$に対応する $X_k(t)$ の中心周波数を基本周波数とする。

$$F_0(t) = \min_{L_F} \mathrm{std}\left(\omega_{\hat{\ell}}/2\pi\right) \tag{3.31}$$

本論文では、$N = 10$、$K/4 \le L_F \le K/2$ とした。探索範囲の上限をパラメータとし、推定された基本周波数のジッターが最小となるときの L_Fを求めることで、倍ピッチと半ピッチによる推定誤差を防ぐことにある。

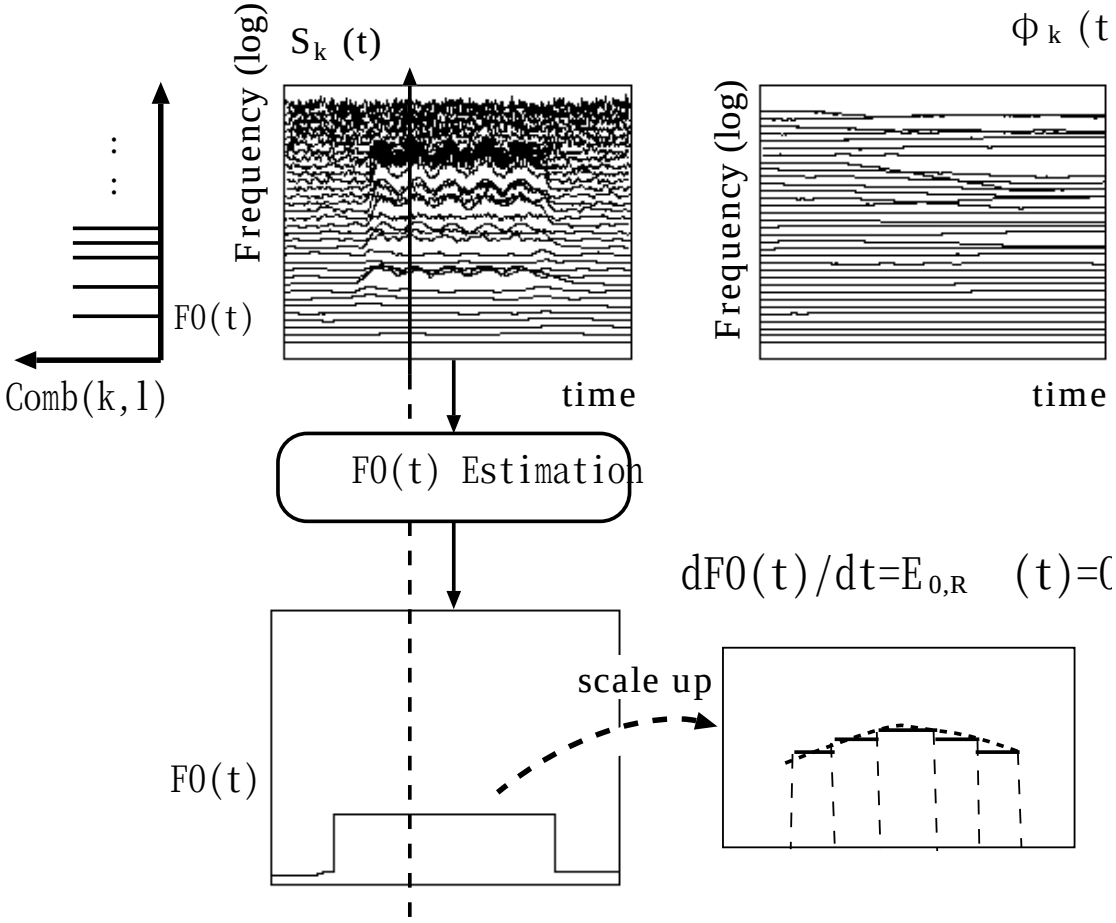

図 3.10: 基本周波数の推定方法

性能評価

上記で構築した基本周波数の推定方法の推定精度を知るために、ATR 音声データベースデータセットにある男性２名（mau, mht）、女性２名（fsu, fkn）の母音 [Takeda *et al.*, 1988] を利用して基本周波数の推定精度を測定した。分離精度については、雑音中でも正確に基本周波数を推定することが望ましいため、SNR を 0〜40 dB まで 10 dB 刻でピンク雑音を加えた 5 種類の混合信号を利用する。また、TEMPO [Kawahara, 97] で推定した原信号の基本周波数を標準パターン $F_0(t)$ とし、雑音が付加された各母音に対し、本方法で推定した基本周波数$\hat{F}_0(t)$ と標準パターン $F_0(t)$ を比較する。ここで、利用する評価尺度は次の三つとした。

1. $F_0(t)$ を信号、$F_0(t) - \hat{F}_0(t)$ を雑音成分と見なした SNR (dB)
2. 最大誤差（$\max(|F_0(t) - \hat{F}_0(t)|)$）(Hz)
3. 平均２乗誤差（root mean square） (Hz)

はじめに、原音声に対する本方法による推定結果を図3.11に示す。この結果では、原音声に対する推定精度はいずれも20 dB以上あり、ほとんど問題なく基本周波数を推定できることがわかる。

次に、妨害雑音（ピンク雑音）をSNRを変化させて原音声に付加した場合の推定精度を図3.12に示す。また、一例として話者mauの母音/u/に対する推定結果を図3.13に示す。これらの結果から、SNRが0 dBと最悪時でも良好に基本周波数を推定できることがわかる。ここでは、比較結果を述べていないが、例えばTEMPOなどを利用した場合では、同条件の下ではほとんど推定できない。

以上の結果から、本方法により雑音にロバストな基本周波数の推定方法を本分析フィルタ群で実現できた。本論文では、以後、特に断わりがない限り、上記の方法で基本周波数を推定する。

3.4.4 グルーピング部

基本周波数の時間変動の制約

一般に、基本周波数は時間的に変動するため、グルーピング部ではこの時間変動（周波数変調）に対応して調波関係や共通の立上り・立下りの関係を考慮しなければならない。そこで、基本周波数の時間変動の制約条件式(3.17)に対し、微小区間で$F_0(t)$は変化しないと考える。つまり、区分的に$dF_0(t)/dt = E_{0,R}(t) = 0$と考える。この場合、基本周波数の時間変化に対する分散量を定義すると、この分散量がある範囲内にあるときの区間を微小区間と解釈できる。従って、この微小区間は

$$\frac{1}{t_h - t_{h-1}} \int_{t_{h-1}}^{t_h} \left| F_0(t) - \overline{F_0(t)} \right|^2 dt \leq (\Delta F_0)^2 \tag{3.32}$$

を用いることで決定できる。但し、微小区間は$t_h - t_{h-1}$であり、$\overline{F_0(t)}$は$F_0(t)$の時間平均、$(\Delta F_0)^2$は分散量の上限である。図3.14に、基本周波数$F_0(t)$と式(3.32)の関係を示す。

一方、本分析フィルタ群で推定された基本周波数$F_0(t)$は、各分析フィルタ群の中心周波数値を取るため、$F_0(t)$は時間的に階段状に変化する。そこで、$F_0(t)$が変化しない区間において、式(3.17)の$E_{0,R}(t) = 0$と解釈すれば、上記の微小区間の考えと同様に区分的に$F_0(t)$が一定であるという考えを利用できる。このとき、$\Delta F_0 = 0$として式(3.32)を利用すれば、容易に微小区間を求めることができる。ここで、階段状に変化する基本周波数$F_0(t)$の不連続点がH個あると仮定し、その点の時刻を$T_1, T_2, \cdots, T_{H-1}, T_H$とする。

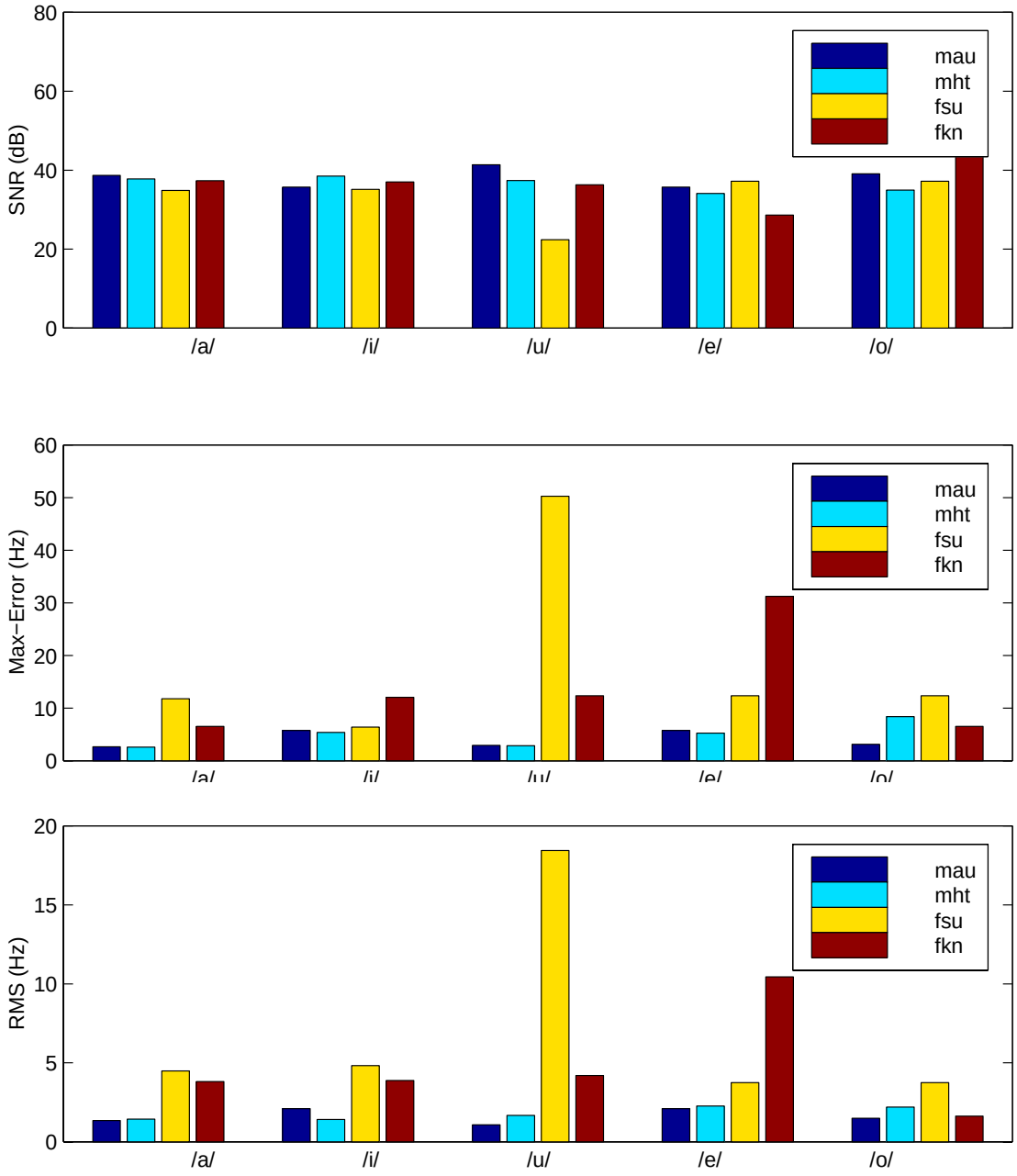

図 3.11: 本推定方法における基本周波数の評価：（上）SNR (dB)、（中）最大誤差 (Hz)、（下）平均２乗誤差 (Hz)

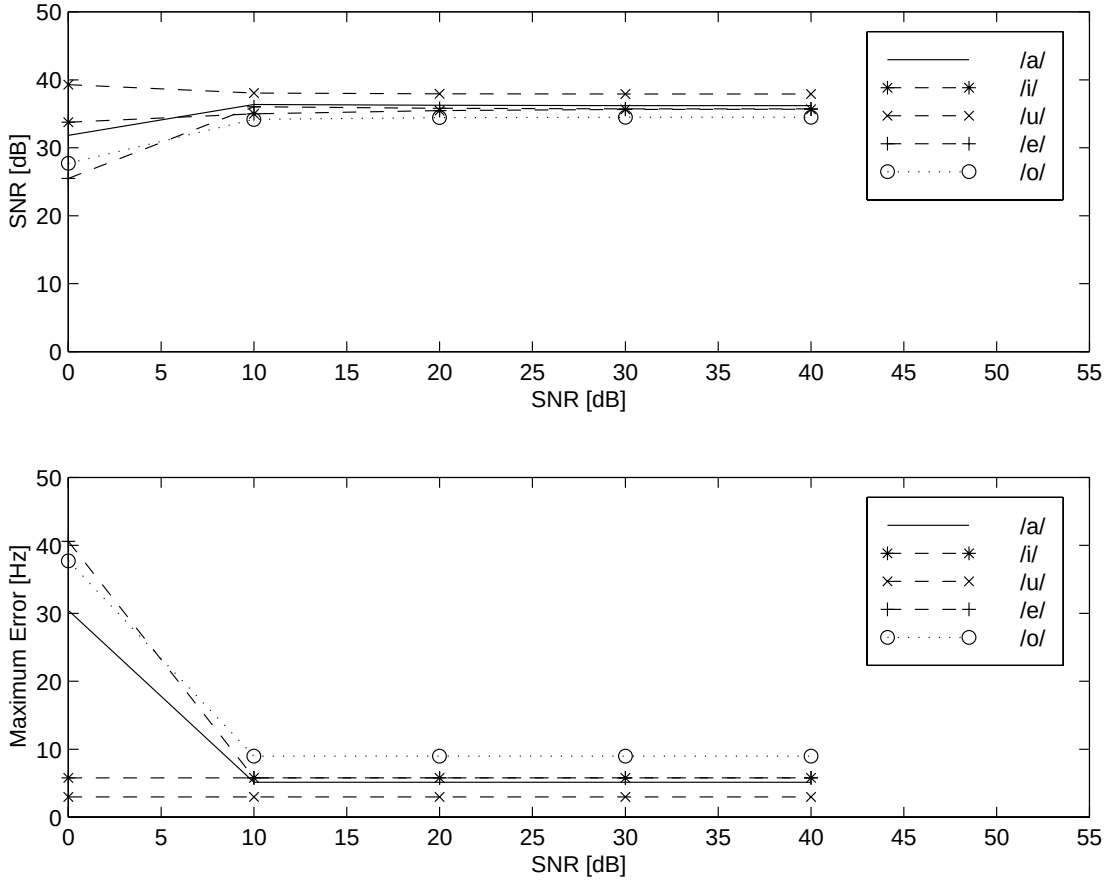

図 3.12: 話者 mau の母音に対する基本周波数の推定のロバスト性：(上）SNR,（下）最大誤差

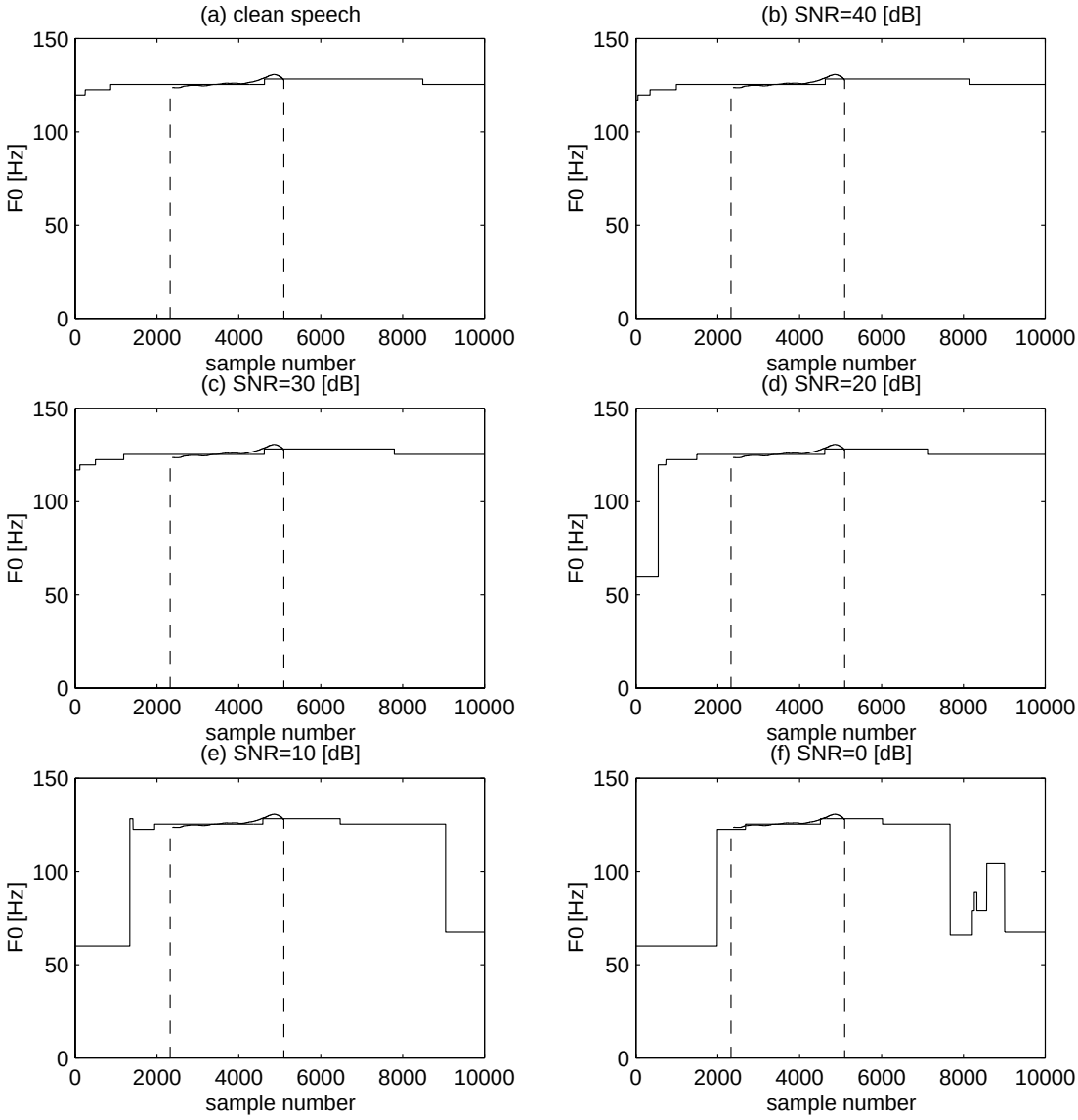

図 3.13: 話者 mau の母音/u/に対する基本周波数の推定結果。図中の実線は本方法によって推定された基本周波数、破線は TEMPO で推定された基本周波数を示す。

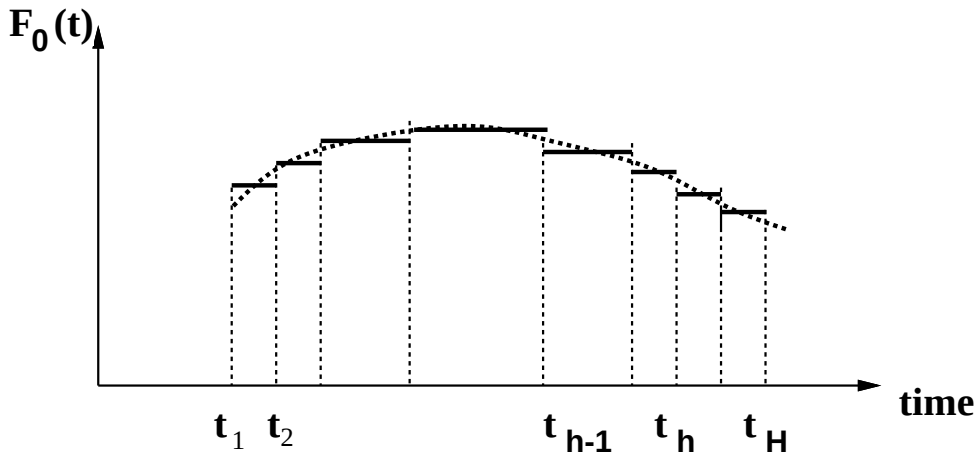

図 3.14: 基本周波数の時間変動

グルーピングの制約条件の実装

各区間で一定となる基本周波数 $F_0(t)$ に対し、制約条件 1 の立上り・立下りの同期性と制約条件 4 の調波関係を実装する。

次に、式 (3.13) と式 (3.14) から、基本波の立上りと立下りをそれぞれ、$T_{\rm S}$ と $T_{\rm E}$ とする。また、一致範囲をそれぞれ$\Delta T_{\rm S} = 50$ ms と$\Delta T_{\rm E} = 100$ ms とする。$\Delta T_{\rm S}$ については、立上りの同期に関する聴取実験の結果を参考にしたものである [柏野, 田中, 1994a]。尚、$X_k(t)$ における $f_1(t)$ の高調波成分の立上り $T_{k,\rm on}$ と立下り $T_{k,\rm off}$ は、それぞれ

(a) 立上り時刻 $T_{k,\rm on}$: $dS_k(t)/dt$ の極大点近傍（±0.25s）にある $|d\phi_k(t)/dt|$ の極大点

(b) 立下り時刻 $T_{k,\rm off}$: $dS_k(t)/dt$ の極小点近傍（±0.25s）にある $|d\phi_k(t)/dt|$ の極大点

で求める。

次に、式 (3.20) から、調波関係にある信号成分が存在する分析フィルタのチャネル番号を

$$\ell = \frac{K}{2} - \left\lceil \frac{\log(n \cdot F_0/f_0)}{\log \alpha} \right\rceil, \qquad n = 1, 2, \cdots, N_{F_0} \tag{3.33}$$

で決定する。但し、αは wavelet 変換のスケールパラメータであり、$\lceil \cdot \rceil$ は、正の無限大方向へ最も近い整数値への丸め記号である。

3.4.5 波形分離部の実装

波形分離部では、二波形分離の対象になる分析フィルタ出力において、瞬時振幅 $S_k(t)$ と瞬時出力位相 $\phi_k(t)$ から、四つのパラメータ ($A_k(t)$, $B_k(t)$, $\theta_{1k}(t)$, $\theta_{2k}(t)$) を決定する。本論文で

は、$C_{k,R}(t)$ と $D_{k,R}(t)$ の係数推定の計算量を抑えるために、式(3.15) を $dA_k(t)/dt = C_{k,1}(t)$、式 (3.16) を $d\theta_{1k}(t)/dt = D_{k,1}(t)$ と仮定し、図 3.5に示す手順でこれらの係数を求める。上記の仮定の場合、$A_k(t)$ と $\theta_{1k}(t)$ は2次の区分多項式で表現できる範囲内で時間変動を許されたことになる。図 3.5の (a)〜(f) の処理については、次節で詳細を説明する。

Kalman filter を用いた推定範囲の決定

はじめに、Kalman filter を用いて $C_{k,0}(t)$ と $D_{k,0}(t)$ を推定する。ここで、$C_k(t) = \int C_{k,0}(t)dt$、$D_k(t) = \int D_{k,0}(t)dt$ とする。推定区間は基本周波数 $F_0(t)$ が一定となる一区間 $[T_{h-1}, T_h]$ である。この区間を、離散時刻 $t_m = T_{h-1} + m/f_s, m = 0, 1, \cdots, (T_h - T_{h-1})f_s$に分割し、時刻 t_mの係数 $C_{k,0}(t_m)$ と $D_{k,0}(t_m)$ の時間変化を

$$
\begin{aligned}
C_k(t_{m+1}) &= C_{k,0}(t_m)\Delta C_k(t_m) + w_m && (3.34)\\
\Delta C_k(t_m) &= 1 + \frac{C_k(t_m) - C_k(t_{m-1})}{C_k(t_m)} && (3.35)\\
D_k(t_{m+1}) &= D_{k,0}(t_m)\Delta D_k(t_m) + w_m && (3.36)\\
\Delta D_k(t_m) &= 1 + \frac{D_k(t_m) - D_k(t_{m-1})}{D_k(t_m)} && (3.37)
\end{aligned}
$$

とする。但し、$t_0 = T_{h-1}$、$t_M = T_h$である。ここで、w_mは、平均 0 で分散 σ_w^2の白色雑音である。

次に、式 (3.34) と式 (3.3)、式 (3.36) と式 (3.3) を $S_k(t)$ で正規化した式を、Kalman filtering 問題 [西山, 中野, 1993]：

$$
\begin{aligned}
\mathbf{x}_{m+1} &= \mathbf{F}_m\mathbf{x}_m + \mathbf{G}_m\mathbf{w}_m \quad \text{（状態方程式）} && (3.38)\\
\mathbf{y}_m &= \mathbf{H}_m\mathbf{x}_m + \mathbf{v}_m \quad \text{（観測方程式）} && (3.39)
\end{aligned}
$$

に対応させる。このとき、上式の各変数は、表3.3のように対応づけられる。次に、式(3.38) と式(3.39) に、Kalman filtering のアルゴリズム [西山, 中野, 1993] を逐次適用すると、最小分散推定量 $\hat{\mathbf{x}}(t_m) = \hat{\mathbf{x}}_{m|m}$と誤差の共分散行列 $\hat{\mathbf{e}}(t_m) = \hat{\mathbf{\Sigma}}_{m|m}$を得る。ここで、推定値と推定誤差をそれぞれ、

$$
\begin{aligned}
\hat{C}_{k,0}(t) &= |d\hat{\mathbf{x}}(t)/dt| && (3.40)\\
P_k(t) &= |d\hat{\mathbf{e}}(t)/dt| && (3.41)\\
\hat{D}_{k,0}(t) &= \arg(d\hat{\mathbf{x}}(t)/dt) && (3.42)\\
Q_k(t) &= \arg(d\hat{\mathbf{e}}(t)/dt) && (3.43)
\end{aligned}
$$

とする。

表 3.3: Kalman filtering の記号の定義.

記号	$C_{k,0}(t)$ の推定	$D_{k,0}(t)$ の推定
観測信号 $\mathbf{y}_m$	$X_k(t_m)$	$\exp(j\phi_k(t_m))$
状態変数 $\mathbf{x}_m$	$C_k(t_m)$	$\exp(jD_k(t_m))$
観測雑音 $\mathbf{v}_m$	$X_{2,k}(t_m)$	$X_{2,k}(t_m)/S_k(t_m)$
システム雑音 $\mathbf{w}_m$	w_m	w_m
状態遷移行列 $\mathbf{F}_m$	$\Delta C_k(t_m)$	$\Delta D_k(t_m)$
観測行列 $\mathbf{H}_m$	$\exp(j\omega_k t_m)$	$\hat{C}_k(t_m)/S_k(t_m)$
駆動行列 $\mathbf{G}_m$	1	1

Spline 補間を用いた候補選定

制約条件 3 におけるなめらかさの制約条件式 (3.18) を満たす $A_k(t)$ と制約条件式 (3.19) を満たす$\theta_{1k}(t)$ を求めるために、$C_{k,1}(t)$ と $D_{k,1}(t)$ の候補を選定する。ここで、式 (3.15) を満たす $C_{k,R}(t),\ R=1$ と, 式 (3.16) を満たす $D_{k,R}(t),\ R=1$ を推定することは、閉区間 $[t_a, t_b]$ において、$A_k^{(R+1)}(\tau_i)=A_{k,i},\ \theta_{1k}^{(R+1)}(\tau_i)=\theta_{1k,i},\ i=1,2,\cdots,I$となる I個の点を通る最もなめらかな補間関数 $A_k^{(R+1)}(t),\ \theta_{1k}^{(R+1)}(t),\ R=1$ を求めることに等しい。この制約での最良補間関数は、$(2R+1)$ 次 Spline 関数であり、唯一存在する [桜井, 1981]。そこで、推定誤差範囲内で Spline 補間された $C_{k,1}(t)$ と $D_{k,1}(t)$ の各候補を求め、その候補から正しい解を一つ求めることで、最もなめらかな瞬時振幅 $A_k(t)$ と瞬時位相 $\theta_{1k}(t)$ の真の解を一意に求める。

本論文では、$R=1$ から、３次 Spline 関数を用いて補間した。また、補間範囲は、$t_a = T_{h-1}$、$t_b = T_h$であり、補間間隔を$\Delta\tau = 15\times(2\pi/\omega_k)/f_s$とした。従って、補間点数 Iは、$I=\lceil(t_b-t_a)/\Delta\tau\rceil$ である。

相関を手がかりにしたパラメータの決定

制約条件 5 の式 (3.21) を用いて、Spline 補間された $C_{k,1}(t)$ の候補を一つの最適解に絞り込む。これは、振幅包絡 $A_k(t)$ 間の相関が、推定誤差内で最大となるときの $C_{k,1}(t)$ を選択することで実現される。

$$\hat{C}_{k,1} = \underset{\hat{C}_{k,0}-P_k\le C_{k,1}\le\hat{C}_{k,0}+P_k}{\arg\max} \frac{<\hat{A}_k, \hat{\hat{A}}_k>}{||\hat{A}_k||||\hat{\hat{A}}_k||} \tag{3.44}$$

但し、$<\cdot>$は内積記号である。ここで、$\hat{A}_k(t)$ は、ある $D_{k,0}(t)$ と Spline 補間された $C_{k,1}(t)$ により得られた振幅包絡であり、$\hat{\hat{A}}_k(t)$ は、

$$\hat{\hat{A}}_k(t) = \frac{1}{N_{F_0}} \sum_{\ell \in \mathbf{L}, \ell \neq k} \frac{\hat{A}_\ell(t)}{\|\hat{A}_\ell(t)\|} \tag{3.45}$$

である。但し、$\mathbf{L}$ は式 (3.20) を満たす ℓの集合である。

次に、Spline 補間された $D_{k,1}(t)$ の候補を一つに絞り込む。これは、上記の手順と同様に、振幅包絡 $A_k(t)$ 間の相関を手がかりに、

$$\hat{D}_{k,1} = \underset{\hat{D}_{k,0}-Q_k \leq D_{k,1} \leq \hat{D}_{k,0}+Q_k}{\arg\max} \frac{<\hat{A}_k, \hat{\hat{A}}_k>}{||\hat{A}_k||||\hat{\hat{A}}_k||} \tag{3.46}$$

で、$D_{k,1}(t)$ の最適解を決定する。但し、$\hat{A}_k(t)$ は$\hat{C}_{k,1}(t)$ と Spline 補間された $D_{k,1}(t)$ により決定された振幅包絡であり、$\hat{\hat{A}}_k(t)$ は式 (3.45) である。

3.5 二波形分離問題の解法の一例

はじめに、AM 調波複合音とピンク帯域雑音が付加された混合信号の分離結果を示す。例えば、図 3.15 (a) の AM 調波複合音に SNR= 10 dB のピンク帯域雑音が付加されたとき、混合信号は図 3.15 (b) となり、分析フィルタ群の出力から図 3.15(c) に示すように基本周波数が推定され、本章で考案した解法により AM 調波複合音が図 3.15 (d) に示すように分離抽出される。この時、図 3.15 (a) に示す原信号を信号音、図 3.15 (d) に示した分離抽出音と原信号の差を雑音と見なしたときの SNR を評価したところ、約 19.1 dB であった。尚、図 3.4は、この混合信号を処理したときの詳細を示している。

ここで、本論文における二波形分離問題の解法の十分性として、原信号と分離抽出された信号の差異が定量的に少ないとき、つまり妨害音が取り除かれ、定量的に差異が改善されたときに、「分離抽出可能である」あるいは「分離抽出できた」と表現する。このとき、二波形分離問題で利用した制約条件は十分条件となる。このとき、先の結果では、原信号に 10 dB の雑音が付加され、分離抽出された信号と原信号の SNR が 19.1 dB であった。従って、分離抽出された信号は、9.1 dB の分離効果（改善）が見られる。

この結果、本章で提案した二波形分離問題の解法は、雑音中から目的の AM 調波複合音を十分に分離抽出できたことがわかる。次章以降では、この二波形分離問題の解法を利用し、発展的構成法に従って二波形分離問題における制約条件の検証を行う。

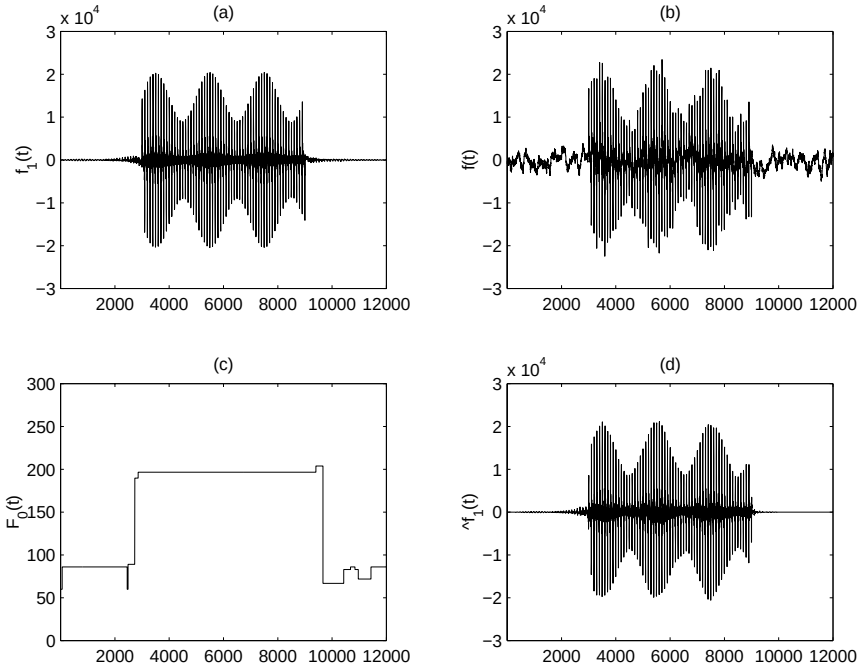

図 3.15: 分離例（AM 調波複合音＋ピンク帯域雑音）：(a) 原信号/a/ $f_1(t)$、(b) 混合信号 $f(t)$、(c) 推定された基本周波数 $F_0(t)$、(d) 分離抽出された信号 $\hat{f}_1(t)$

3.6 むすび

本章では、二波形分離問題の解法を提案した。

はじめに、前章で定義した AM–FM 調波複合音を分離対象として取り扱える二波形分離問題を定式化した。このとき、定式化した問題において、観測された混合信号の瞬時振幅と瞬時出力位相から四つのパラメータ（二波形の瞬時振幅と瞬時入力位相）を一意に解けないことを示した。この結果から、本論文で取り扱う二波形分離問題が不良設定の逆問題であることを示した。

次に、この不良設定問題を一意に解くために利用する制約条件を定式化した。これは、前章で示した四つの発見的規則の取り扱い方とそれに対する音の構成に深く関係するものであった。

次に前章で述べた二波形分離における物理量（振幅と位相、基本周波数）の求め方を示した。また、聴覚特性を考慮できる分析フィルタ群は定 Q gammatone filterbank で実装され、基本周波数はフィルタ群の出力における Comb filtering から得られた。

最後に、二波形分離問題を解くためのアルゴリズムを実装した。特に、実装後の処理の確認として、各ブロックにおける処理例として、分析合成フィルタ群の変換・逆変換の結果

（フィルタバンクを素通りさせた結果）と基本周波数推定部のロバスト性の評価結果を示した。ここで、本論文における「制約条件の十分性」を「定量的に分離効果があること」と定義した。その後で、本章で提案した二波形分離問題の解法が、十分性の意味で AM–FM 調波複合音を分離抽出できることを示した。

以上結果から、本章で提案した二波形分離問題の解法により、AM–FM 調波複合音を分離抽出することが可能である。次章以降では、音の分離抽出における聴覚の計算理論を構築するために、発展的構築法を展開し、二波形分離問題における制約条件の検証を行う。

第 4 章

二波形分離モデルの検証

4.1 まえがき

本章では、第３章で提案した二波形分離問題の解法における物理量と制約条件の十分条件と有効性の検証を目的とする。そこで、発展的構築法の手順に従い、AM–FM 調波複合音を利用して二波形分離問題の解法による分離精度を評価することで、制約条件の十分性と有効性を検証する。

はじめに、分離抽出の対象となる音を AM 単一成分音とした場合について、二波形分離問題の解法による分離精度を評価する。ここでは、単一成分の瞬時振幅 $A_k(t)$ に対し、漸近的変化の多項式近似となめらかさの制約条件を検証する。主に、瞬時振幅の時間変動に対する区分多項式近似の表現精度について検証する。

次に、分離抽出の対象となる音を AM–FM 調波複合音とした場合について、二波形分離問題の解法による分離精度を評価する。特に、基本周波数が一定のものと時間的に変動するもの、つまり周波数変調された場合とそうでない場合の二種類の AM–FM 調波複合音を利用する。また、雑音をランダム帯域雑音およびピンク帯域雑音とする。この検証では、主に瞬時位相の時間変動に対する区分多項式近似の表現精度について分離精度を評価する。また、同時に複合成分における瞬時振幅と基本周波数の時間変動による影響も検証する。その後で、分離対象の音が周波数変調されているかどうかで分離精度の違いを比較する。

最後に、二波形分離問題の解法で利用した制約条件の有効性を検証する。これは、AM–FM 調波複合音に対し、制約条件を一つずつ省略した場合の分離精度を評価することで、制約条件の有効性を検証する。

4.2 二波形分離モデルの検証手順

二波形分離モデルで利用した制約条件の十分性と有効性を示すためには、次の項目を検証する必要がある。

1. 瞬時振幅 $A_k(t)$ と瞬時位相 $\theta_{1k}(t)$ を拘束する制約条件の十分性
   (a) 瞬時振幅 $A_k(t)$ に対する漸近的変化の効果
   (b) 瞬時位相 $\theta_{1k}(t)$ に対する漸近的変化の効果
   (c) 基本周波数 $F_0(t)$ の時間変化に対する漸近的変化の効果
2. 四つの発見的規則に対応した制約条件の有効性
   (a) 共通の立上り・立下りに関する制約条件の有効性

(b) $A_k(t)$ と$\theta_{1k}(t)$ を拘束する制約条件の有効性

(c) 振幅包絡間の相関に関する制約条件の有効性

(d) 調波関係に関する制約条件の有効性

はじめに項目１について説明する。項目１ (a),(b) では、単一成分の $A_k(t)$、$\theta_{1k}(t)$ に対する制約条件の十分性を、項目１ (c) では、複合成分に渡る $A_k(t)$ および $\theta_{1k}(t)$ に対する基本周波数の影響を調べることが目的である。しかし、$\theta_{1k}(t)$ は、基本周波数の時間変動に影響を受けるため、検証項目として独立に評価することは困難である。そこで、分離抽出の対象音として AM 単一成分音を利用し、単一成分における $A_k(t)$ および $f_1(t)$ の分離精度を評価することで、項目１ (a) を検証する。その後で、分離抽出の対象となる音として AM–FM 調波複合音を利用し、複合成分における $A_k(t)$ および $\theta_{1k}(t)$ の分離精度を評価することで、項目１ (b) および項目１ (c) を検証する。このとき、項目１ (a) に関しては複合成分間における $A_k(t)$ の検証も可能である。また、AM–FM 調波複合音において、周波数変調の有無の条件、つまり AM 調波複合音と AM–FM 調波複合音の二つの信号に対して評価することで、基本周波数に対する制約条件の十分性も検証できる。

最後に、項目２について説明する。項目２では、制約条件の有効性を検証することを目的にしているため、分離抽出の対象となる音として AM–FM 調波複合音を利用し、この二波形分離問題において利用する制約条件を一つずつ省略した場合の分離精度の評価を行う。この評価により、四つの発見的規則それぞれを利用することの有効性が確認できる。

以上の手順で検証シミュレーションを行うことにより、二波形分離問題で利用した制約条件の十分性および有効性を示すことができる。

4.2.1 分離精度の評価

本論文では、二波形分離モデルの分離精度を評価するために、次に述べる二種類の評価尺度を用意した。

一つは、瞬時振幅 $A_k(t)$ の分離精度の時間平均である。この評価尺度を用いる目的は、信号と雑音が同一周波数領域に存在（二つの信号成分の和が $S_k(t)$ に混在）しても正確に $A_k(t)$ を分離できるか調べることである。この評価尺度を Precision と呼び、次式で定義する。

$$\frac{1}{T}\int_0^T \left(10\log_{10}\frac{\sum_{k=1}^{K}\tilde{A}_k(t)^2}{\sum_{k=1}^{K}(\tilde{A}_k(t)-A_k(t))^2}\right)dt \qquad \text{(dB)} \tag{4.1}$$

ここで、$\tilde{A}_k(t)$ はあらかじめ $f_1(t)$ を分析フィルタ群に展開して得られた瞬時振幅であり、$A_k(t)$ は本方法によって分離抽出された$\hat{f}_1(t)$ の瞬時振幅である。

もう一つは、$f_1(t)$ と $\hat{f}_1(t)$ の差を雑音とみなした時間領域における $f_1(t)$ の SNR である。この評価尺度を用いる目的は、二波形の位相も正確に分離でき、かつ正確に波形レベルで復元できるかを調べることである。この評価尺度を Segregation accuracy と呼び、次式で定義する。

$$10\log_{10}\frac{\int_0^T f_1(t)^2dt}{\int_0^T (f_1(t)-\hat{f}_1(t))^2dt} \qquad (\mathrm{dB}) \tag{4.2}$$

以後、特に断わりがない限り、本論文では上記二つの評価尺度を利用する。

4.3 AM 単一成分音を利用した制約条件の十分性の検証

4.3.1 検証シミュレーションにおける二波形分離問題の仮定

ここでは、二波形分離問題における $f_1(t)$ を AM 単一成分音、$f_2(t)$ を妨害雑音とする。このとき、AM 単一成分音は周波数変調されていないため、時間的にいつでも同じ周波数成分をもつことになる。そこで、AM 単一成分音の中心周波数とそれを解析するために利用する分析フィルタの中心周波数 $\omega_k/2\pi$が完全に一致するものと仮定する。これは強い制約のように思われるが、周波数変調されていない単一周波数成分音であるため、分離抽出したい信号の瞬時入力位相 $\theta_{1k}(t)$ を扱い易くなる。そこで、本章では、$D_{k,1}(t)=0$ でかつ $\theta_{1k}(t)=0$ と仮定する。

さて、上記の二波形分離問題の場合、第２章で述べたように Bregman によって提唱された四つの発見的規則すべてを利用する必要なない。つまり、必要と考えられる制約条件は、少なくとも AM 単一成分音の振幅の時間変化を拘束することである。第３章で考案したアルゴリズムでは、振幅包絡間の相関を手がかりに最適解を一意に導いたが、この状況では、AM 単一成分音の振幅包絡を他の振幅包絡と相関を取ることができない。そこで、本問題に限り、AM 単一成分音とこれを妨害する雑音を振幅変調された帯域雑音とした二波形分離問題とする。この二波形分離問題は、共変調マスキング解除として知られる心理現象を想定した問題設定に相当する。この仮定の下では、AM 単一成分音を雑音の振幅包絡間の変動の一致の有無に合わせて、分離抽出できるかどうかを検討することができる。

次に、$A_k(t)$ に対する漸近的変化に関する制約条件の十分性を検証するために、$A_k(t)$ の時間変化に対する区分多項式近似 $C_{k,R}(t)$ の表現精度を評価する必要がある。そこで、第３章で提案した解法の $C_{k,R}(t)=C_{k,1}(t)$ 以外に、$C_{k,R}(t)=C_{k,0}(t)$, $C_{k,R}(t)=0$ の三つの場合について分離精度を評価し、漸近的変化に関する制約条件の十分性を検証する。

次に、上記三つの場合に対応した分離精度を評価するため、第３章で提案した方法におけるパラメータ決定法を説明する。

(a) $\theta_{1k}(t) = 0$ と仮定する。

(b) Kalman filter を用いて、式 (3.15) の $C_{k,0}(t)$ を推定する。但し、$\hat{C}_{k,0}(t)$ は最小分散推定値、$P_k(t)$ は推定誤差を示す。

(c) 推定誤差内 $\hat{C}_{k,0}(t) - P_k(t) \leq C_{k,1}(t) \leq \hat{C}_{k,0}(t) + P_k(t)$ から、Spline 補間された $C_{k,1}(t)$ の候補を求める。

(d) 式 (4.4) の相関値最大を尺度に、$\hat{C}_{k,1}(t)$ を決定する。

(e) $\hat{C}_{k,1}(t)$ から $\theta_{2k}(t) = \theta_k(t)$ を求める。

図 4.1: $C_{k,R}(t) = C_{k,1}(t)$ の場合のパラメータ決定手順

4.3.2 二波形分離問題の解法におけるパラメータ決定法の変更点

$C_{k,R}(t) = C_{k,1}(t)$ および $C_{k,0}(t)$ の場合のパラメータ決定法

はじめに、$C_{k,R}(t) = C_{k,1}(t)$ の場合のパラメータ決定法を述べる。そこで、第 3 章の制約条件 5 で定義された振幅包絡間の変動の一致に関する数理工学的な制約条件を次式で再定義する。

制約条件 5 - 2 (振幅包絡 $B_k(t)$ 間の相関) 振幅包絡 $B_k(t)$ は隣接する分析フィルタにおける振幅包絡 $B_\ell(t)$ に強い相関がなければならない：

$$\frac{B_k(t)}{\|B_k(t)\|} \approx \frac{B_{k\pm\ell}(t)}{\|B_{k\pm\ell}(t)\|}, \qquad \ell = 1, 2, \cdots, L \tag{4.3}$$

但し、$\|\cdot\|$ はノルム記号である。

■

ここでは、制約条件 5 の再定義により、波形分離部の相関値最大の尺度の部分だけを変更する。次に、再実装された波形分離部のパラメータ決定手順を図 4.1に示す。

ここで、制約条件 5-2 を規範に、雑音の振幅包絡間の相関が最大になるときの $C_{k,1}(t)$ を求める。これは、次式で実現できる。

$$\hat{C}_{k,1} = \underset{\hat{C}_{k,0} - P_k \leq C_{k,1} \leq \hat{C}_{k,0} + P_k}{\arg\max} \frac{< \hat{B}_k, \hat{\hat{B}}_k >}{||\hat{B}_k|| ||\hat{\hat{B}}_k||} \tag{4.4}$$

但し、

$$\hat{\hat{B}}_k(t) = \frac{1}{2L} \sum_{\ell=-L, \ell\neq 0}^{L} \frac{\hat{B}_{k+\ell}(t)}{\|\hat{B}_{k+\ell}(t)\|} \tag{4.5}$$

(a) $\theta_{1k}(t) = 0$ と仮定する。

(b) 対象となる微小区間において、式 (4.9) の$\underline{C_{k,0}} \leq C_{k,0} \leq \overline{C_{k,0}}$を求める。

(c) 式 (3.22) から各 $C_{k,0}$に対する入力位相差 $\hat{\theta}_k(t)$ を求め、式 (3.7) と式 (3.8) から二波形の瞬時振幅 $A_k(t)$ と $B_k(t)$ を求める。

(d) 隣接する分析フィルタ特性から、$A_k(t)$ の通過成分を $A_{k\pm\ell}(t)$ を求める

(e) 瞬時振幅 $S_{k\pm\ell}(t)$, 瞬時出力位相 $\phi_{k\pm\ell}(t)$ および瞬時振幅 $A_{k\pm\ell}(t)$ から、入力位相差 $\theta_{k\pm\ell}(t)$ を求める。

(f) 式 (3.8) から瞬時振幅 $\hat{B}_{k\pm\ell}(t)$ を求める。

(g) 式 (4.10) の振幅包絡 $B_k(t)$ 間の相関値最大を尺度に $C_{k,0}$の最適解を求める。

図 4.2: $C_{k,R}(t) = 0$ の場合のパラメータ決定手順

である。ここで、L は隣接する分析フィルタの参照数を示す。特に指定しない限り、本論文では $L = 1$ とする。

次に、$C_{k,R}(t) = C_{k,0}(t)$ とした場合のパラメータ決定法を述べる。これは、図 4.1において、$\hat{C}_{k,1}(t)$ の代わりに Kalman filter を用いて推定された$\hat{C}_{k,0}(t)$ を直接利用する方法に対応する。従って、図 4.1 (c), (d) を省略すればよい。

$C_{k,R}(t) = 0$ の場合のパラメータ決定法

最後に、$C_{k,R}(t) = 0$ つまり、区分的に $dA_k(t)/dt = C_{k,R}(t) = 0$ と仮定した場合のパラメータ決定法を述べる。これは、$C_{k,R}(t)$ の係数推定の計算量を最も軽減した方法に対応する。上記の多項式近似の設定は、区分的に $dA_k(t)/dt = 0$ から、$A_k(t)$ は区分的に定数である、つまり、区分的に $A_k(t) = C_{k,0}$（0 次近似）で表現することを意味する。そのため、$A_k(t) = C_{k,0}$を表現する各区分間の不連続点を拘束する必要がある。そこで、分離区間を I 個の微小区間 $\Delta t = M/f_0$に分割し、この分割された各微小区間に対し、波形分離を行う方法を考える。この処理を図 4.2に示めす。また、詳細については以下で説明する。

まず、この微小区間 Δt の接合境界を拘束するために、発見的規則 (ii) の漸近的変化（なめらかさ）を利用する。本章では、“分離を行った微小区間（$T_r - \Delta t \leq t < T_r$）と分離を行う微小区間（$T_r \leq t < T_r + \Delta t$）の境界 T_rにおいて、各パラメータが連続性を保持しなければならない” と解釈する。この定性的な規則を次の数理工学的な制約条件として表記する。これは制約条件 3 を再定義することに相当する。

制約条件 3 - 2 (漸近的変化（時間的近接）) 時間領域 $T_r \leq t < T_r + \Delta t$ において分離を行うとき、二波形の瞬時振幅 $A_k(t)$, $B_k(t)$ と入力位相差 $\theta_k(t)$ は、分離境界（$t = T_r$）の前後において、ある幅 ΔA, ΔB, $\Delta\theta$ 以内で接合されていなければならない：

$$|A_k(T_r + 0) - A_k(T_r - 0)| \leq \Delta A \tag{4.6}$$

$$|B_k(T_r + 0) - B_k(T_r - 0)| \leq \Delta B \tag{4.7}$$

$$|\theta_k(T_r + 0) - \theta_k(T_r - 0)| \leq \Delta\theta \tag{4.8}$$

■

式 (3.7)、式 (3.8)、式 (3.22) から、$A_k(t)$ と $B_k(t)$ および $\theta_k(t)$ が未定係数 $C_{k,0}$の関数となっていることがわかる。この点に着目すれば、制約条件 3-2 は、ある境界 T_rにおける連続性を保持した形で $C_{k,0}$の取り得る範囲を

$$\underline{C_{k,0}} \leq C_{k,0} \leq \overline{C_{k,0}} \tag{4.9}$$

に限定することと解釈できる。但し、$\underline{C_{k,0}}$と$\overline{C_{k,0}}$は、この境界における未定係数 $C_{k,0}$の上限と下限である。このことから、各微小区間毎に上記の推定範囲を狭めることで最適解の探索範囲を狭めることができる。

次に、狭められた探索範囲内から、一意な解を求める。式 (3.8) と式 (3.22) から $B_k(t)$ は $C_{k,0}$の関数であることがわかる。そこで、ある $C_{k,0}$により決定された振幅包絡を$\hat{B}_k(t)$ とおく。ここで、制約条件 5-2 を規範に、振幅包絡間の相関が最大になるときの $C_{k,0}$を次式で求める。

$$\hat{C}_{k,0} = \underset{\underline{C_{k,0}} \leq C_{k,0} \leq \overline{C_{k,0}}}{\arg\max} \frac{\langle \hat{B}_k, \hat{\hat{B}}_k \rangle}{||\hat{B}_k|| ||\hat{\hat{B}}_k||} \tag{4.10}$$

以上の最適解導出の計算を、各微小区間毎に繰り返し、最適な $C_{k,0}$を求めることで、一意な入力位相差 $\theta_k(t)$ を求める。最後に、一意な $\theta_k(t)$ から、二波形の瞬時振幅 $A_k(t)$ と $B_k(t)$ を求める。

ここでは、微小区間を$\Delta t = 3/f_0$、隣接する分析フィルタ数を $L = 1$ とした。また、分離区間（$T_{k,\mathrm{on}} \leq t \leq T_{k,\mathrm{off}}$）において、$S_k(t)$ の最大値を $S_{\max}$としたとき、$\Delta B = 0.027 S_{\max}$、$\Delta\theta = \pi/20$ と一定値にしたが、ΔA は $C_k(t) = C_{k,0}, T_r \leq t < T_r + \Delta t$ の影響により一定値にすることが困難であるため、式 (4.6) に基づき$\Delta A = |A_k(T_r - \Delta t) - A_k(T_r - 2\Delta t)|$ とした。

4.3.3 シミュレーションデータ

検証シミュレーションで利用する実験データとして、分離抽出音 $f_1(t)$ を次に示す三種類の AM 単一成分音、妨害雑音 $f_2(t)$ を振幅変調されたランダム帯域雑音とする。ここで、$f_{11}(t)$ を純音、$f_{12}(t)$ をランプ関数で振幅変調された単一成分音、$f_{13}(t)$ を正弦波信号で振幅変調された単一成分音とした。これを図 4.3に示す。

$$f_{11}(t) = F_{\mathrm{BP}}(g_{11}), \quad \text{（変調なし）} \tag{4.11}$$

$$g_{11}(t) = \begin{cases} 1200\sin(2\pi f_0 t), & 0.3 \le t \le 0.7 \\ 0, & \text{otherwise} \end{cases}$$

$$f_{12}(t) = F_{\mathrm{BP}}(g_{12}), \quad \text{（ランプ関数の変調）} \tag{4.12}$$

$$g_{12}(t) = \begin{cases} 2000\left(1 + t - \frac{3}{10}\right)\sin(2\pi f_0 t), & 0.3 \le t \le 0.7 \\ 0, & \text{otherwise} \end{cases}$$

$$f_{13}(t) = F_{\mathrm{BP}}(g_{13}), \quad \text{（正弦波の変調）} \tag{4.13}$$

$$g_{13}(t) = \begin{cases} 2000\left(1 + \frac{1}{10}\sin(2\pi f_c t)\right)\sin(2\pi f_0 t), & 0.3 \le t \le 0.7 \\ 0, & \text{otherwise} \end{cases}$$

但し、$F_{\mathrm{BP}}(\cdot)$ は中心周波数が f_0で帯域幅が 23 Hz の帯域通過フィルタを表し、$f_0 = 600$ Hz、$f_c = 10$ Hz とする。ここで、ランダム帯域雑音は、60〜6000 Hz に帯域制限された白色雑音である。また、これに 30 Hz の低域通過フィルタをかけたもので振幅変調されたランダム帯域雑音である。尚、ここでは過変調を起こさないように振幅値にバイアス値を足して作成した。このとき、$f_2(t)$ の帯域幅は 1 kHz とし、$f_{11}(t)$ と $f_2(t)$ の SN 比は−8.5 dB とした。

4.3.4 検証シミュレーションの条件

混合信号は、SNR を−10〜20 dB まで 5 dB 刻に、振幅変調されたランダム帯域雑音を単一成分音に付加して作成した。また、7 個の各混合信号に対して Precision および Segregation accuracy の二種類の尺度で分離精度を評価する。

シミュレーション条件として、

- Condition 1: $C_{k,R}(t) = C_{k,1}(t)$ の場合

  図 4.1の導出方法を利用して $A_k(t)$ を決定

- Condition 2: $C_{k,R}(t) = C_{k,0}(t)$ の場合

  Kalman filter で推定した $C_{k,0}(t)$ を利用して $A_k(t)$ を決定

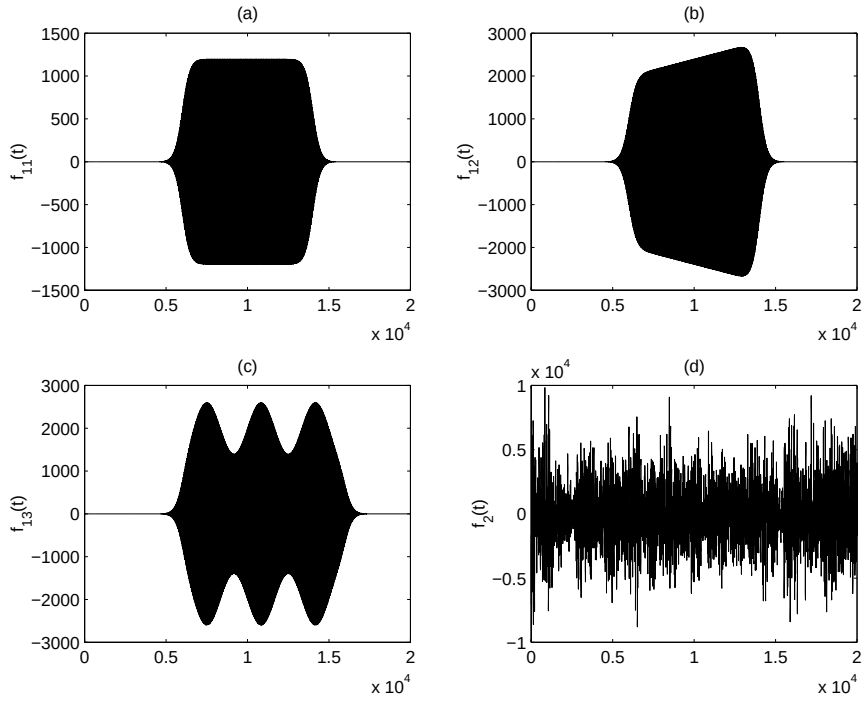

図 4.3: 実験データ：(a) 純音, (b) ランプ信号, (c) 正弦波振幅をもつ単一音, (d) 振幅変調されたランダム帯域雑音.

- Condition 3: $C_{k,R}(t) = 0$ の場合

  図 4.2の最適解導出の方法を利用して $A_k(t)$ を決定

- No processing:

  何も処理をせず、分析合成系を通過させたもの。$A_k(t)$ の評価では $A_k(t)$ を $S_k(t)$ で代用し、$f_1(t)$ の評価では $f_1(t)$ を $f(t)$ で代用する。

上記四つの比較条件において、先の図 4.3に示した三つの AM 単一成分音の分離抽出の結果を検証する。

4.3.5 検証結果

はじめに、純音に振幅変調されたランダム帯域雑音が付加された場合の分離精度を測定した。この結果を図 4.4に示す。図 4.4 (a) は Precision（原信号の瞬時振幅を信号音、原信号と分離抽出した信号のそれぞれの瞬時振幅の差を雑音と見なしたときの SNR）を、図 4.4 (b) は Segregation accuracy（波形レベルでの原信号と分離抽出した信号の SNR）を示す。例えば、図 4.7 (a) の純音 $f_{11}(t)$ に、SNR= 0 dB の振幅変調されたランダム帯域雑音が付加

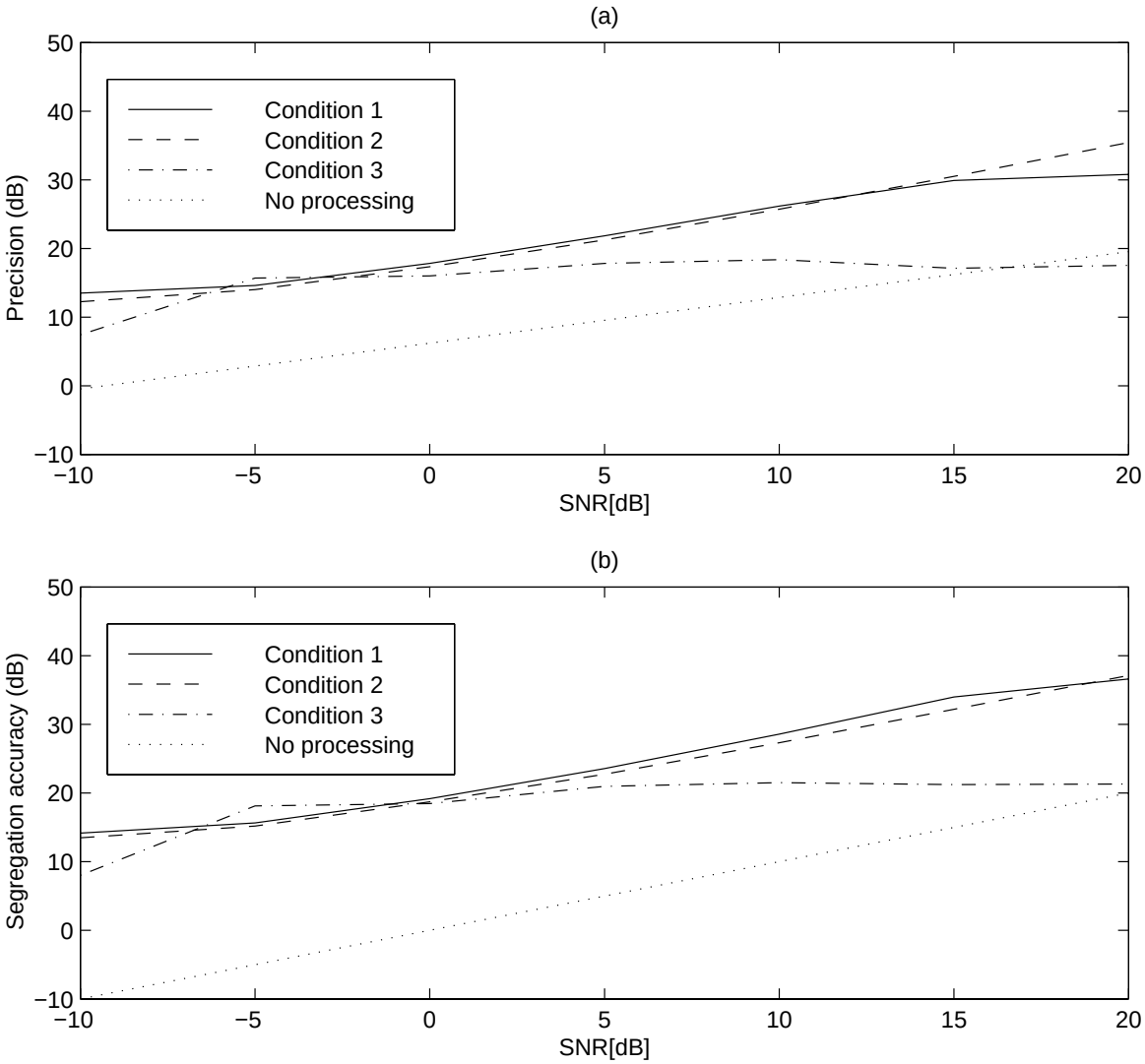

図 4.4: 純音に振幅変調されたランダム帯域雑音が付加された場合の SN 比と分離精度の関係：(a) Precision, (b) Segregation accuracy

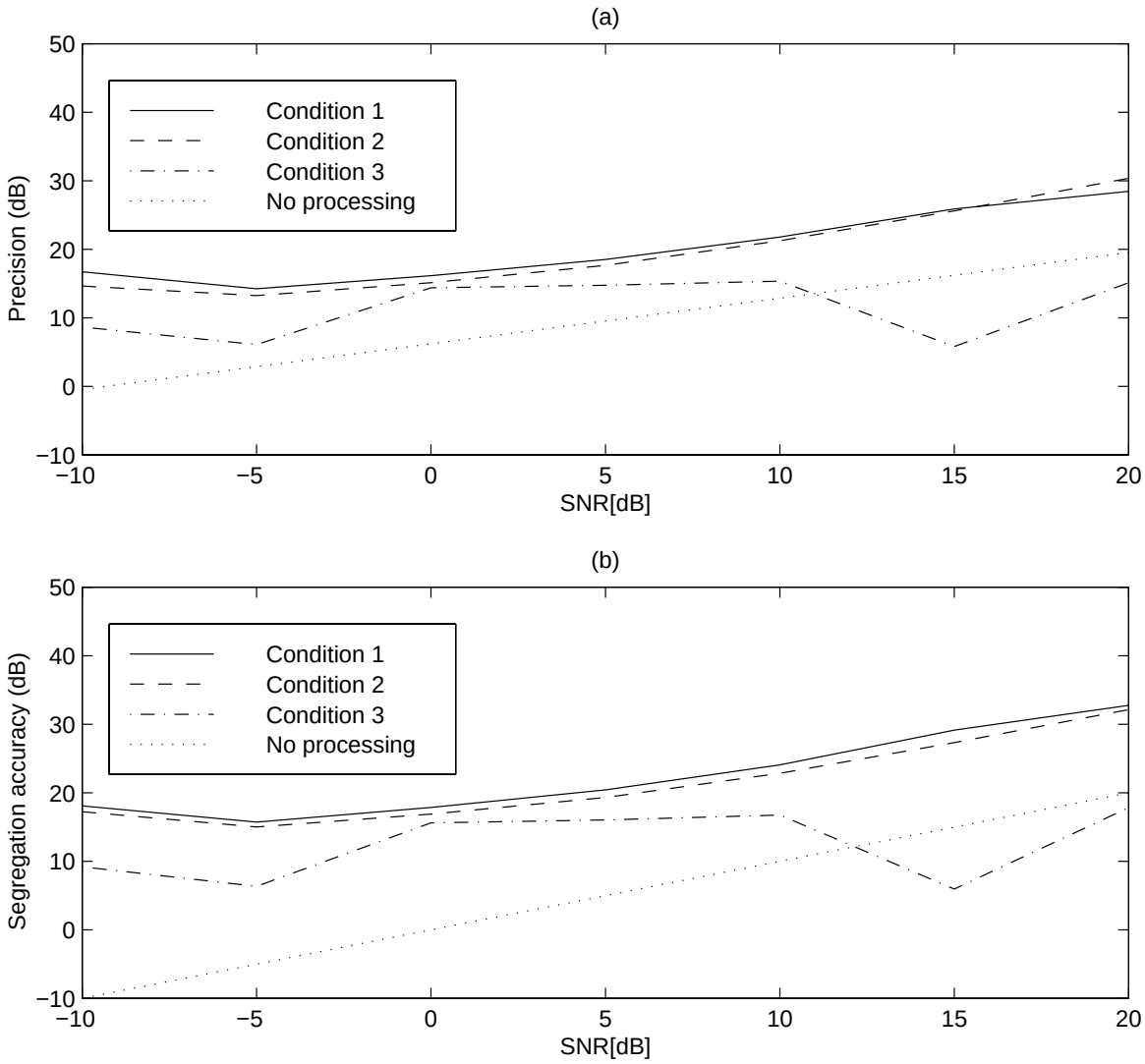

図 4.5: AM 単一成分音（ランプ関数）に振幅変調されたランダム帯域雑音が付加された場合の SN 比と分離精度の関係：(a) Precision, (b) Segregation accuracy

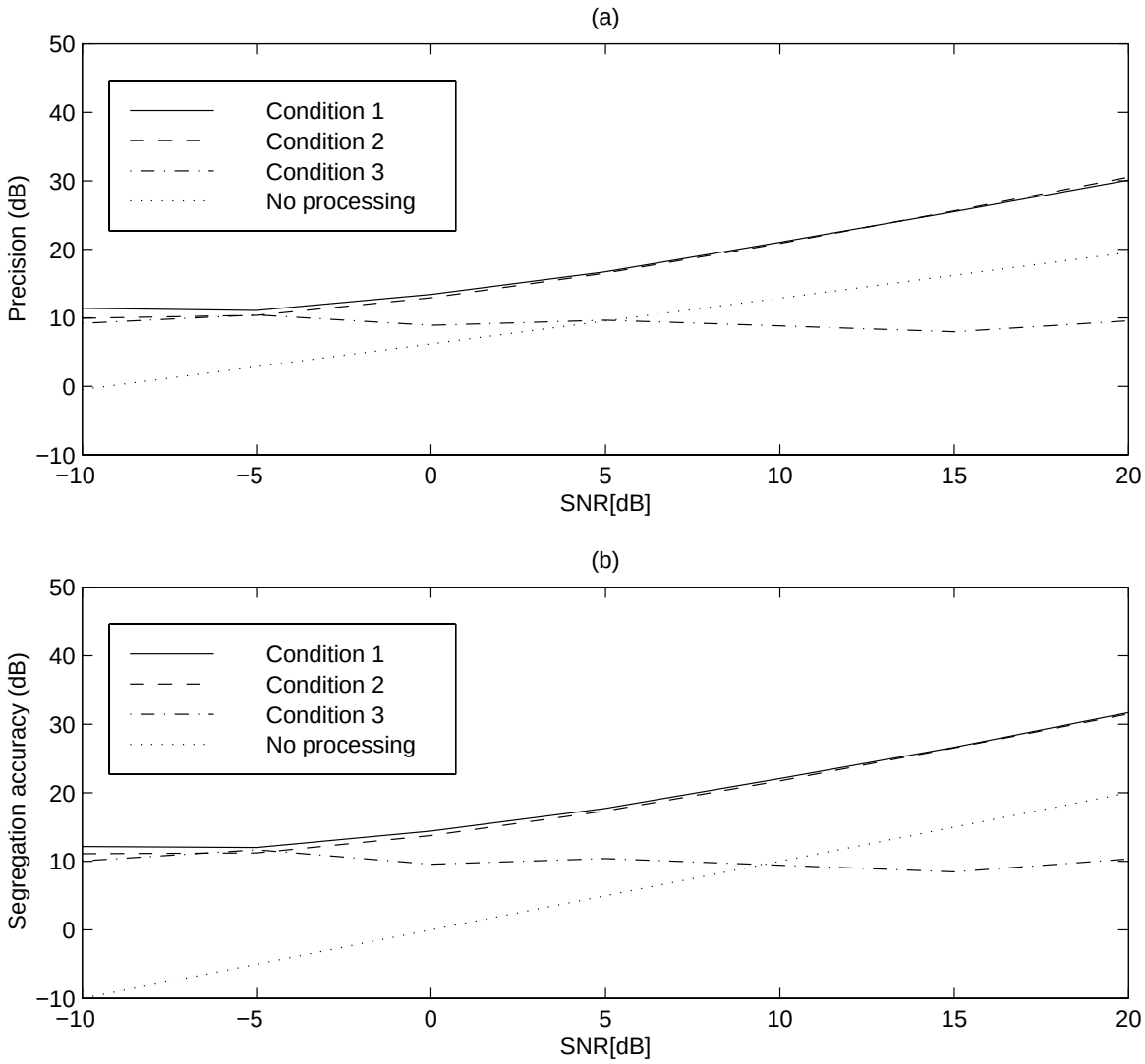

図 4.6: AM 単一成分音（正弦波の変調）に振幅変調されたランダム帯域雑音が付加された場合の SN 比と分離精度の関係：(a) Precision, (b) Segregation accuracy

された混合信号（図 4.7 (b)）の場合、$C_{k,1}(t)$ を用いた二波形分離アルゴリズム（Condition 1）による分離結果は、図 4.7 (c), (d) のようになる。ここで、図 4.7 (c) は、分離抽出された純音の振幅包絡 $A_k(t)$ を、図 4.7 (d) は、分離抽出された純音を示す。このとき、二波形分離アルゴリズムは、16.5 dB の Segregation accuracy で純音を分離抽出できた。

次に、ランプ関数で振幅変調された単一成分音に振幅変調されたランダム帯域雑音が付加された場合の分離精度を測定した。この結果を図 4.5に示す。図中の (a),(b) は図 4.4に示したものと同様のものを示す。例えば、図 4.8 (a) のランプ関数で振幅変調された単一成分音 $f_{12}(t)$ に、SNR= 0 dB の振幅変調されたランダム帯域雑音が付加された混合信号（図 4.8 (b)）の場合、$C_{k,1}(t)$ を用いた二波形分離アルゴリズム（Condition 1）による分離結果は、図 4.8 (c), (d) のようになる。ここで、図 4.8 (c) は、分離抽出された純音の振幅包絡 $A_k(t)$ を、図 4.8 (d) は、分離抽出された $f_{12}(t)$ を示す。このとき、二波形分離アルゴリズムは、17.0 dB の Segregation accuracy で AM 単一成分音を分離抽出できた。

最後に、正弦波信号で振幅変調された単一成分音に振幅変調されたランダム帯域雑音が付加された場合の分離精度を測定した。この結果を図 4.6に示す。図中の (a),(b) は図 4.4に示したものと同様のものを示す。例えば、図 4.9 (a) の正弦波で振幅変調された単一成分音 $f_{13}(t)$ に、SNR= 0 dB の振幅変調されたランダム帯域雑音が付加された混合信号（図 4.9 (b)）の場合、$C_{k,1}(t)$ を用いた二波形分離アルゴリズム（Condition 1）による分離結果は、図 4.9 (c), (d) のようになる。ここで、図 4.9 (c) は、分離抽出された純音の振幅包絡 $A_k(t)$ を、図 4.9 (d) は、分離抽出された $f_{13}(t)$ を示す。このとき、二波形分離アルゴリズムは、12.0 dB の Segregation accuracy で AM 単一成分音を分離抽出できた。

上記すべての分離結果から、図 4.1に示した解法で定量的に分離効果がみられ、一番よい分離精度を示したことがわかる。この結果から、AM 単一成分音を分離抽出するための十分条件は、漸近的変化（多項式近似）で仮定した $C_{k,R}(t)$ を $R = 1$ として問題を解くことである。また、振幅変調された単一成分音のとき、漸近的変化のなめらかさを時間的近接（境界条件）として利用するよりも Spline 補間によるなめらかさの規範を利用することの有効性が顕著に現れた。

4.3.6 純音に対する分離抽出の考察

図 4.3 (a) に示した純音の立上り時刻を 12.5 msec ずつ移動させて作った 10 個の混合信号に対する二波形分離を行った。ここで、図 4.3の純音に対し、図 4.2に示した方法を用いて二波形分離を行った。ここで、10 個の混合信号に対する Segregation accuracy の平均値を求めたところ、$\hat{f}_1(t)$ が 12.9 dB（標準偏差 2.58）、$\hat{f}_2(t)$ が 10.1 dB（標準偏差 0.20）で

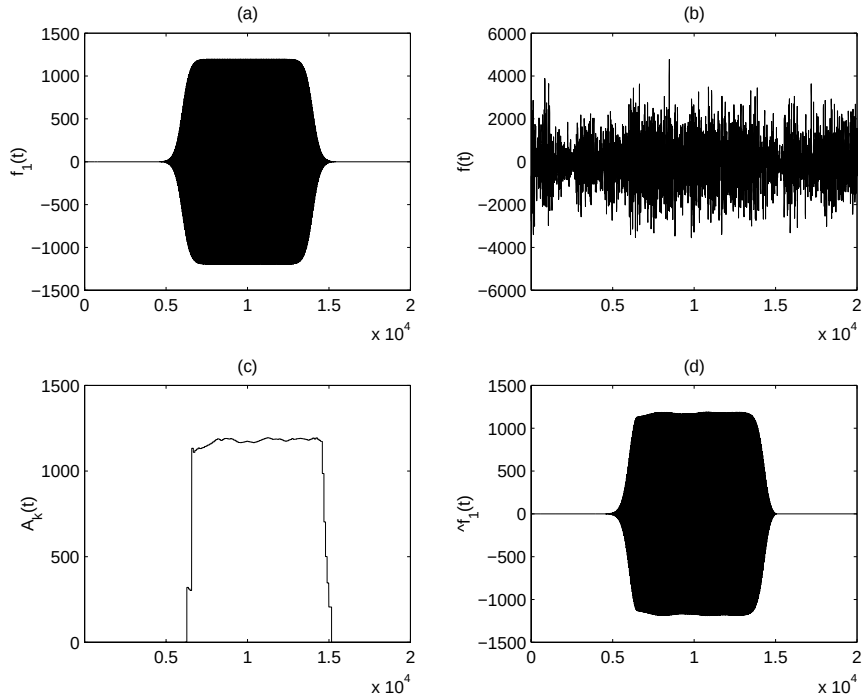

図 4.7: 分離例（Condition 1, 純音の場合）：(a) 原信号 $f_1(t)$, (b) 混合信号 $f(t)$, SNR= 0 dB, (c) 瞬時振幅 $A_k(t)$, (d) 分離抽出した信号$\hat{f}_1(t)$

あった。

以上の結果から、振幅包絡 $A_k(t)$ の時間変動を区分的に定数で拘束し、なめらかさの規範を境界条件に対応した制約条件を利用しても、純音を分離抽出できることがわかった。従って、純音の場合の十分条件は、区分多項式近似を $C_{k,R}(t) = 0$（定数 $C_{k,0}$）として問題を解くことである。

次に下記の二点について本モデルを考察してみる。

帯域通過フィルタを利用した場合との比較

ここでは、数理工学的な方法との比較検討を行う。例えば、中心周波数が 600 Hz、帯域幅が 23 Hz の帯域通過フィルタ（BPF）を考える。これは、分析フィルタ群における一つの分析フィルタに等しい。ここで、この帯域通過フィルタを用いて $f(t)$ の雑音を抑制し、これを$\hat{f}_1(t)$ とするような分離方法を考えてみる。上記の実験と全く同じ条件のもとで、純音に振幅変調されたランダム帯域雑音を付加した混合信号に対し帯域通過フィルタリングを行ったところ、$\hat{f}_1(t)$ の Segregation accuracy は約 8.1 dB となった。一方、本方法では、12.9 dB であった。この結果から、本方法は帯域通過フィルタを単に利用した分離方法と比較しても分離精度が優れていることがわかる。

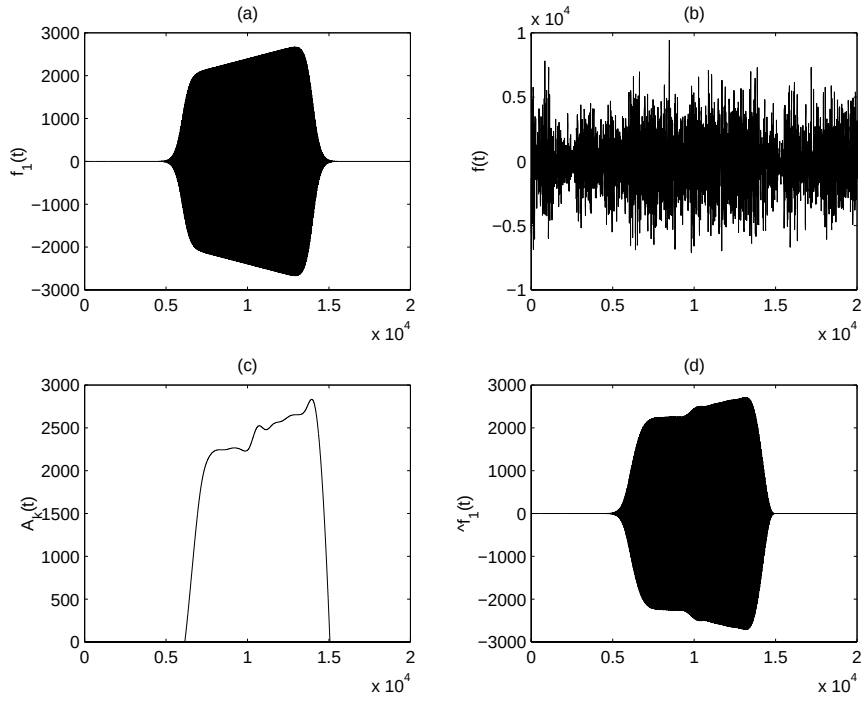

図 4.8: 分離例（Condition 1, ランプ関数で振幅変調された単一成分音の場合）: (a) 原信号 $f_1(t)$, (b) 混合信号 $f(t)$, SNR= 0 dB, (c) 瞬時振幅 $A_k(t)$, (d) 分離抽出した信号$\hat{f}_1(t)$

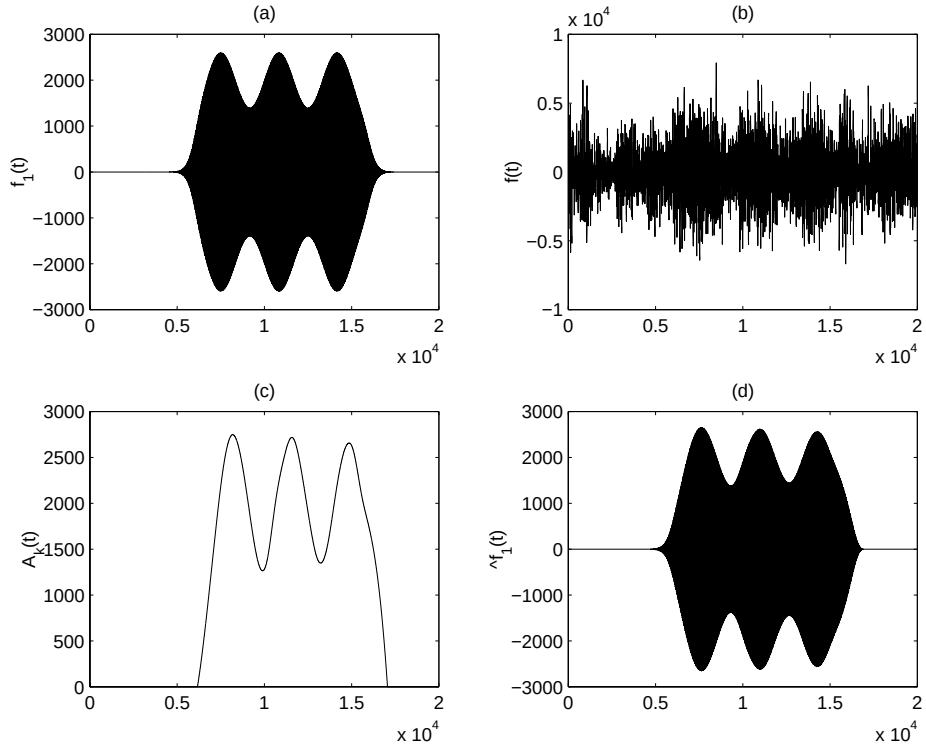

図 4.9: 分離例（Condition 1, 正弦波で振幅変調された単一成分音の場合）: (a) 原信号 $f_1(t)$, (b) 混合信号 $f(t)$, SNR= 0 dB, (c) 瞬時振幅 $A_k(t)$, (d) 分離抽出した信号$\hat{f}_1(t)$

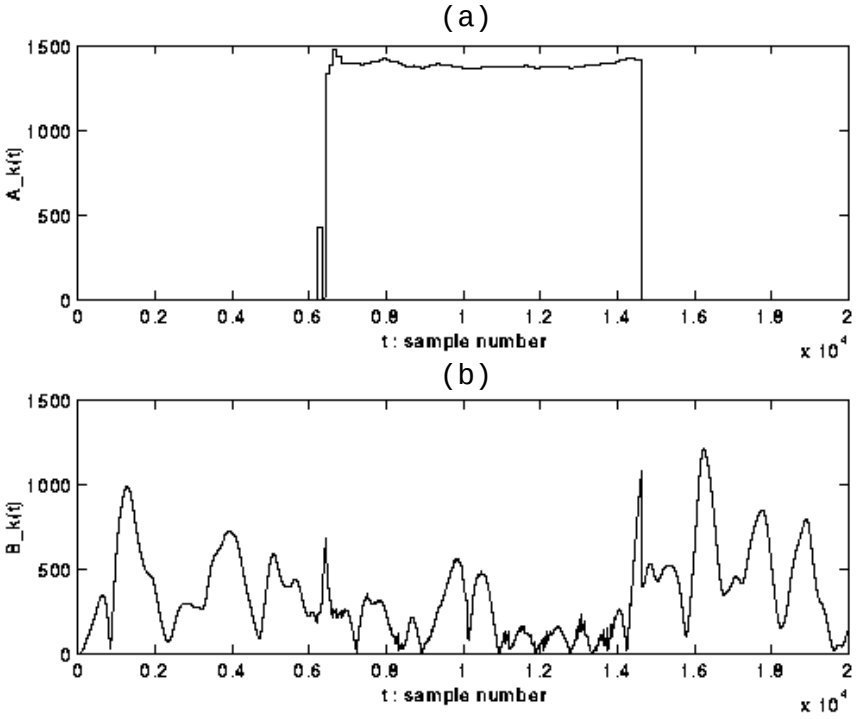

図 4.10: 制約条件 3-2 すべてを利用した場合の分離結果：(a) 瞬時振幅 $A_k(t)$, (b) 瞬時振幅 $B_k(t)$

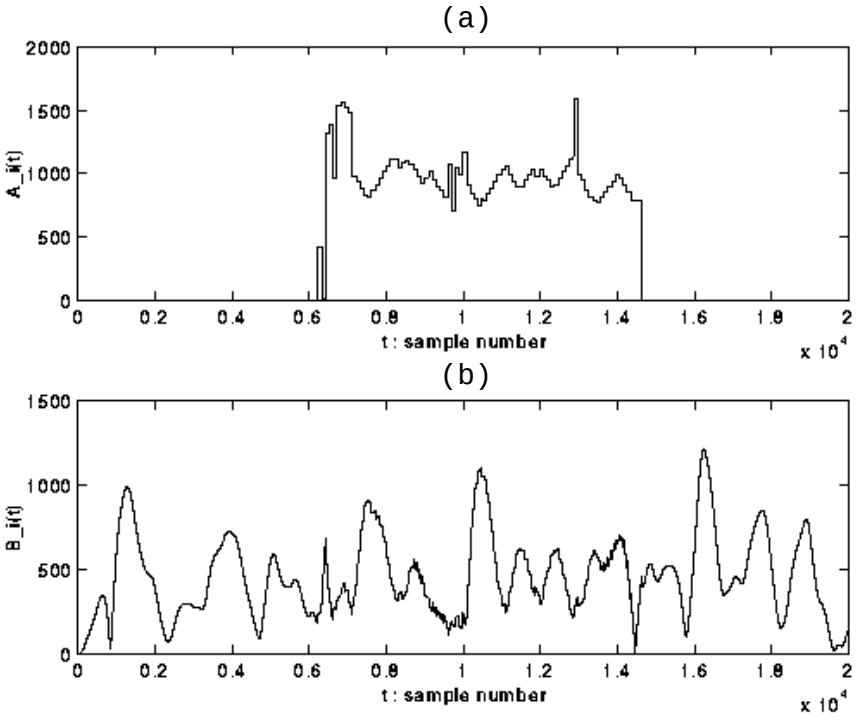

図 4.11: 制約条件 3-2 の一部（$B_k(t)$ に関する制約）を利用した場合の分離結果：(a) 瞬時振幅 $A_k(t)$, (b) 瞬時振幅 $B_k(t)$

振幅包絡 $B_k(t)$ に対する制約条件のみを利用した場合の効果

ここでは、制約条件を更に緩和した場合の効果を検討する。上記の制約条件により、最適解の探索範囲が狭められる。次に、制約条件 3-2 の漸近的変化（時間的近接）において、すべての制約を利用するのではなく、式 (4.7) のみを更に利用した場合を考えてみる。これは、分離抽出したい信号のなめらかさを一切見ずに、雑音の振幅包絡のみを拘束することに等しい。ここで、先の混合信号に対する分離結果を比較した。図 4.10に、制約条件 3-2 すべてを利用した場合の瞬時振幅の分離結果を、図 4.11に、上記の方法を利用した場合の瞬時振幅の分離結果を示す。この結果、雑音の振幅包絡のなめらかさのみを拘束した場合、純音の振幅包絡 $A_k(t)$ が多少変動してしまうことが明らかになった。この比較結果から、純音の振幅包絡にも、なめらかさの制約条件で拘束する必要があることがわかる。

4.4 AM–FM 調波複合音を利用した制約条件の十分性の検証

4.4.1 検証シミュレーションにおける二波形分離モデルの仮定

前節では、分離抽出したい音を、周波数変調されていない AM 単一成分音とした場合について二波形分離問題の解法を検証した。本節では、二波形分離問題における $f_1(t)$ を AM–FM 調波複合音、$f_2(t)$ を妨害音として、$A_k(t)$ と$\theta_{1k}(t)$、基本周波数 $F_0(t)$ に対する制約条件の十分性の検証を行う。

まず、前節では、AM 単一成分音の瞬時振幅と瞬時入力位相の時間変動を考慮したが、瞬時入力位相に限っては信号の周波数と分析フィルタの中心周波数が一致するという仮定をした。しかし、調波複合音を考えた場合、瞬時入力位相の時間変動を十分考慮しなければ、波形レベルで完全に復元することは難しい。また、AM 調波複合音は周波数変調を受けていないため、単純に基本周波数の整数倍を考えればよいが、周波数変調を受けた場合は、時々刻々と変化する基本周波数に対し、調波成分の関係を拘束しなければならい。そこで、二種類の AM–FM 調波複合音に対する検証シミュレーションを考える。一つは、基本周波数が一定な AM 調波複合音であり、もう一つは基本周波数が時間的に変動する（周波数変調）された AM–FM 調波複合音である。また、妨害音はランダム帯域雑音とピンク帯域雑音とする。この場合、二種類の信号の利用により、基本周波数の時間変動の有無による影響および瞬時振幅と瞬時位相の分離精度の評価を行うことができる。具体的には、$\theta_{1k}(t)$ に対する漸近的変化に関する制約条件の十分性を検証するために、$\theta_{1k}(t)$ の時間変化

に対する区分多項式近似 $D_{k,R}(t)$ の表現精度を評価する。そこで、第３章で提案した解法の $D_{k,R}(t) = D_{k,1}(t)$ 以外に、$D_{k,R}(t) = D_{k,0}(t)$、$D_{k,R}(t) = 0$ の三つの場合について分離精度を評価し、漸近的変化に関する制約条件の十分性を検証する。

次に、上記三つの場合に対応した分離精度を評価するために、第３章で提案した解法におけるパラメータ決定法の変更点を説明する。

4.4.2 二波形分離問題の解法におけるパラメータ決定の変更点

本節では、分離抽出する信号の瞬時振幅と瞬時入力位相、および基本周波数の時間変動を区分多項式近似で拘束することで、目的の音を分離抽出するものと仮定する。また、高調波成分を調波関係と立上り・立下りの同期で拘束しなければならない。従って、Bregman によって提唱された四つの発見的規則をすべて利用し、第３章で提案した二波形分離問題の解法をそのまま利用する。そして、二波形分離問題で利用する制約条件について考察する。尚、前節では、制約条件 5 について、雑音の振幅包絡間の変動の一致を考慮したが、ここでは分離抽出したい信号の振幅包絡の変動の一致を利用する。

そのため、$D_{k,R}(t) = D_{k,1}(t)$ の場合のパラメータ決定法には、特に変更点がなく、図 3.5 のアルゴリズムをそのまま利用できる。また、$D_{k,R}(t) = D_{k,0}(t)$ とした場合のパラメータ決定法は、図 3.5中の$\hat{D}_{k,1}(t)$ の代わりに Kalman filter を用いて推定された$\hat{D}_{k,0}(t)$ を直接利用することになる。

$D_{k,R}(t) = 0$ の場合のパラメータ決定法

$D_{k,R}(t) = 0$、つまり区分的に $\theta_{1k}(t) = D_{k,0}$（$d\theta_{1k}(t)/dt = 0$）と仮定した場合のパラメータ決定法を述べる。

ここで、瞬時位相 $\theta_{1k}(t)$ の時間変化を区分的に定数とした場合のパラメータ決定の手順を、図 4.12に示す。0 次の区分多項式の場合は、第３章で述べた方法で $D_{k,1}(t)$ の計算処理を除く方法となる。詳細の決定方法は以下で説明する。

はじめに、$[-\pi/2, \pi/2]$ 内からある $D_{k,0}$の値を選び、各 $D_{k,0}$に対応する最適な $C_{k,1}(t)$ を選定する。そして、次式に示すように、$[-\pi/2, \pi/2]$ で振幅包絡 $A_k(t)$ 間の相関を最大にする $D_{k,0}$を選ぶことで、一意な $D_{k,0}$とそれに対応する $C_{k,1}(t)$ を求める。

$$\hat{D}_{k,0} = \underset{-\pi/2 \le D_{k,0} \le \pi/2}{\arg\max} \frac{< \hat{A}_k, \hat{\hat{A}}_k >}{||\hat{A}_k|| ||\hat{\hat{A}}_k||}. \tag{4.14}$$

但し、$\hat{A}_k(t)$ は$\hat{C}_{1,k}(t)$ で決定された瞬時振幅であり、$\hat{\hat{A}}_k(t)$ は式 (3.45) で決定された瞬時振幅である。上記以外は、これまでに提案した方法と全く同じ方法で二波形分離問題を解く。

(a) $-\pi/2 \leq D_{k,0} \leq \pi/2$ 内の、ある $D_{k,0}$を選択する。

(b) Kalman filter を用いて $C_{k,0}(t)$ を推定する。

(c) 推定誤差内 $\hat{C}_{k,0}(t) - P_k(t) \leq C_{k,1}(t) \leq \hat{C}_{k,0}(t) + P_k(t)$ から、Spline 補間された $C_{k,1}(t)$ の候補を求める。

(d) 振幅包絡間の相関値最大を尺度に、$C_{k,1}(t)$ を求める。

(e) (a)〜(e) を繰り返す。

(f) 式 (4.14) の振幅包絡間の相関値最大を尺度に、$D_{k,0}$を求める。

図 4.12: パラメータの決定手順

4.4.3 シミュレーションデータ

周波数変調されていない AM 調波複合音の実験データとして、$f_1(t)$ を図 4.15 (a) に示す振幅変調された調波複合音とする。このときの振幅変調周波数は 10 Hz の正弦波であり、前章の正弦波信号で振幅変調された単一成分音が 20 個の調波構造で形成されるものである。また、基本周波数は 200 Hz 一定である。この AM 調波複合音に対する妨害雑音 $f_2(t)$ は、60〜6000 Hz で帯域制限された二種類の帯域雑音（ピンク帯域雑音とランダム帯域雑音）とする。$f(t)$ の SNR を 0 dB から 20 dB まで 5 dB 刻みに変化させた５種類の混合信号 $f(t)$ をシミュレーションデータとして利用する。

次に、振幅と周波数の両方で変調されている AM–FM 調波複合音の実験データとして、$f_1(t)$ を図 4.19 (a) に示す LMA（Log Magnitude Approximation）合成母音 [今井, 北村, 1978] とする。また、妨害雑音 $f_2(t)$ として、60〜6000 Hz で帯域制限された二種類の帯域雑音（ランダム帯域雑音とピンク帯域雑音）とする。但し、$f_1(t)$ の基本周波数は平均が 125 Hz、変動幅が 5 Hz（123〜128 Hz）であり、LMA で合成された母音/a/とした。$f(t)$ の SNR を 0 dB から 20 dB まで 5 dB 刻みに変化させた５種類の混合信号 $f(t)$ をシミュレーションデータとして利用する。

4.4.4 検証シミュレーションの条件

シミュレーション条件として、

- Condition 1: $D_{k,R}(t) = D_{k,1}(t)$ の場合

  図 3.5の導出方法を利用して $A_k(t)$ と$\theta_{1k}(t)$ を決定

- Condition 2: $D_{k,R}(t) = D_{k,0}(t)$ の場合

  Kalman filter で推定した $D_{k,0}(t)$ を利用して $A_k(t)$ と$\theta_{1k}(t)$ を決定

- Condition 3: $D_{k,R}(t) = 0$ の場合

  図 4.12の最適解導出の方法を利用して $A_k(t)$ と$\theta_{1k}(t)$ を決定

- No processing:

  何も処理をせず、分析合成系を通過させた場合。$A_k(t)$ の評価では $A_k(t)$ を $S_k(t)$ で代用し、$f_1(t)$ の評価では $f_1(t)$ を $f(t)$ で代用する。

上記四つの比較条件において、AM–FM 調波複合音の分離抽出の結果を検証する。

4.4.5 検証結果

はじめに、AM 調波複合音とピンク帯域雑音が付加された混合信号の分離結果を図 4.13 に示す。図 4.13 (a) は瞬時振幅に対する Precision を、図 4.13 (b) は Segregation accuracy（原信号と分離抽出された信号の差を雑音と見なした SNR）を示す。例えば、図 4.15 (a) の合成母音に SNR= 10 dB のピンク帯域雑音が付加されたとき、混合信号は図 4.15 (b) となり、分析フィルタ群の出力から基本周波数が図 4.15(c) に示すように推定され、第３章で考案された解法により AM 調波複合音が図 4.15 (d) に示すように分離抽出される。ここで、$D_{k,0}$の定数制約とした Condition 3 では、波形レベルまでは完全に復元できず、Segregation accuracy での評価はできなかった。この結果は、残り三つの検証シミュレーションでも確認された。

次に、AM 調波複合音とランダム帯域雑音が付加された混合信号の分離結果を図 4.14に示す。先のシミュレーション結果と同様、二つの評価尺度において、SNR が良好な場合（20 dB）は若干の分離精度の向上しか見られないが、SNR が最悪な場合（0 dB）は顕著な分離精度の向上が見られる。例えば、図 4.16 (a) の AM 調波複合音に SNR= 10 dB のランダム帯域雑音が付加されたとき、混合信号は図 4.16 (b) となり、分析フィルタ群の出力から基本周波数が図 4.16 (c) に示すように推定され、第３章で考案した解法により AM 調波複合音が図 4.16 (d) に示すように分離抽出される。

次に、合成母音とピンク帯域雑音が付加された混合信号の分離結果を図 4.17に示す。図 4.17 (a) は瞬時振幅に対する Precision を、図 4.17 (b) は原信号と分離抽出された信号の差を雑音と見なした SNR を示す。二つの評価尺度において、SNR が良好な場合（20 dB）は若干の分離精度の向上しか見られないが、SNR が最悪な場合（0 dB）は顕著に分離精度の

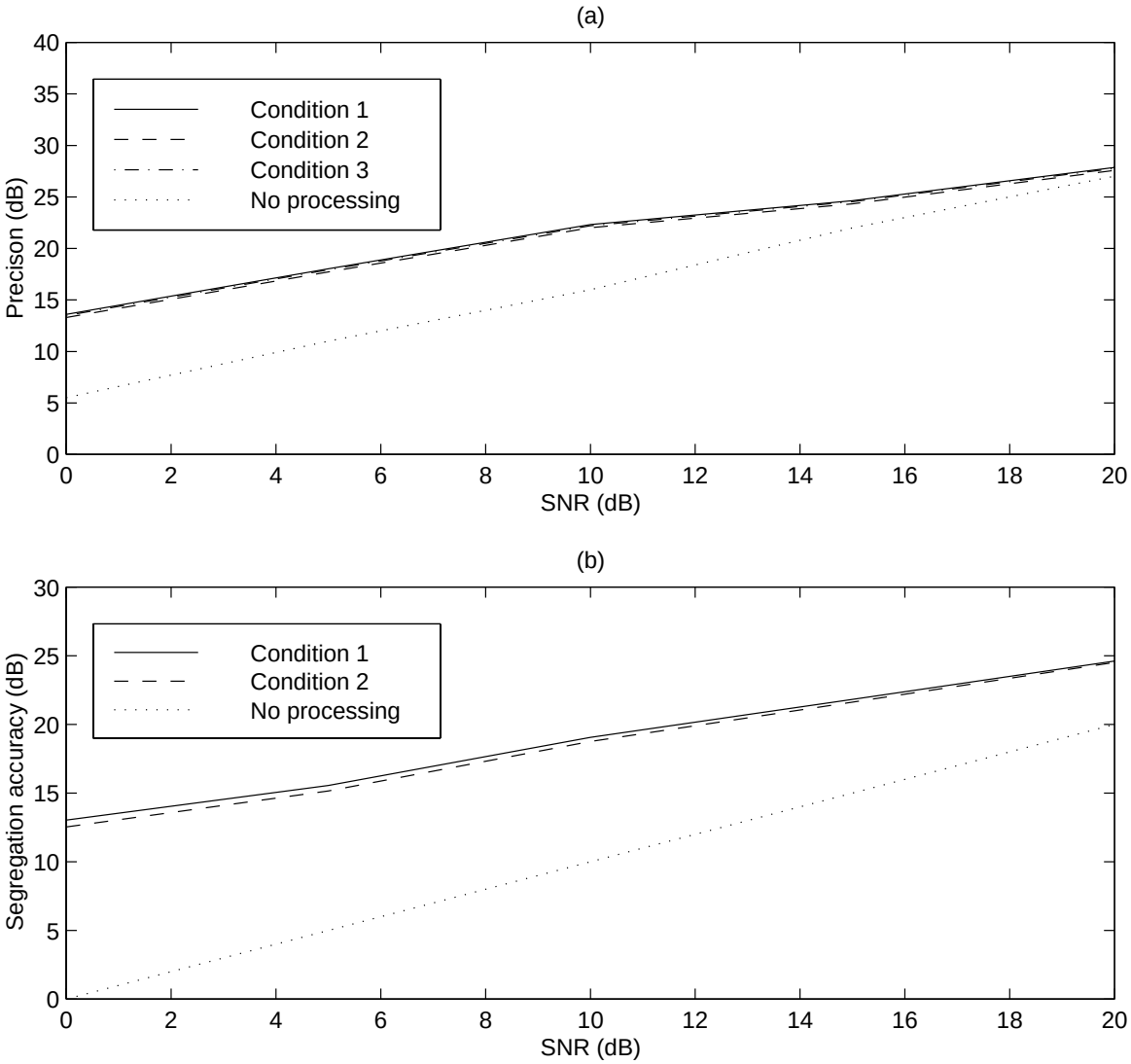

図 4.13: 分離精度の比較（AM 調波複合音＋ピンク帯域雑音の場合）: (a) Precision, (b) Segregation accuracy

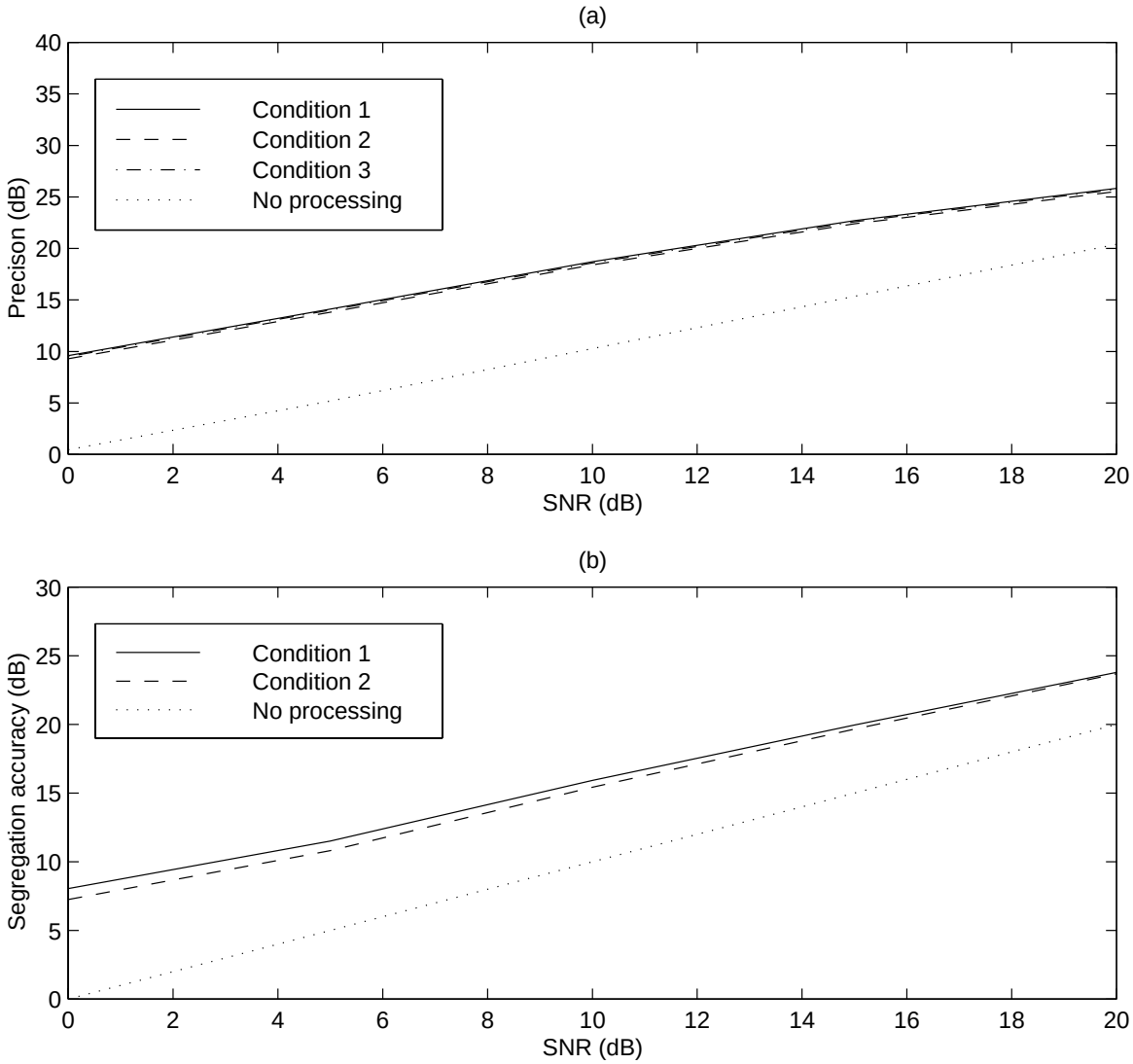

図 4.14: 分離精度の比較（AM 調波複合音＋ランダム帯域雑音の場合）：(a) Precision, (b) Segregation accuracy

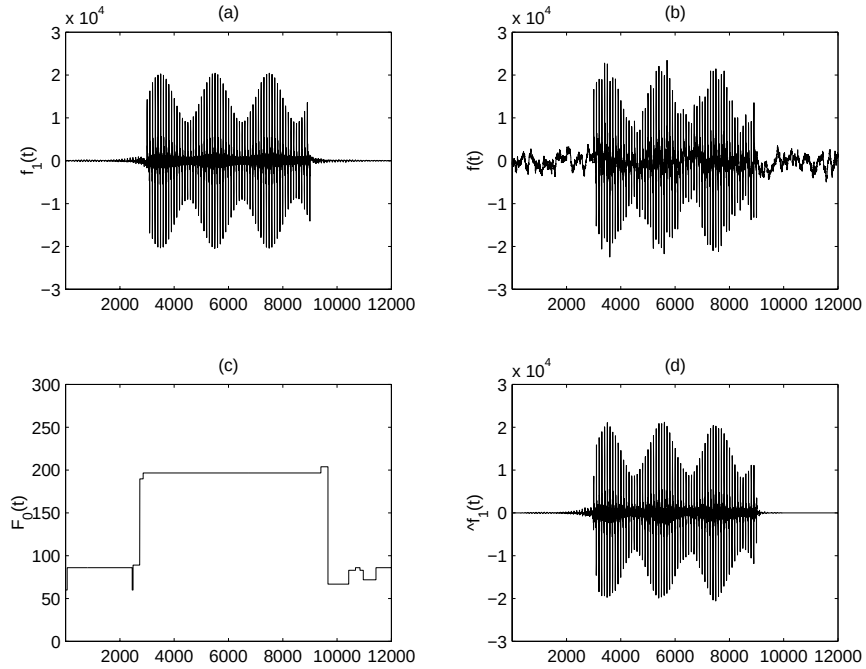

図 4.15: 分離例（Condition 1, AM 調波複合音＋ピンク帯域雑音）：(a) 原信号/a/ $f_1(t)$, (b) 混合信号 $f(t)$, SNR= 10 dB, (c) 推定された基本周波数 $F_0(t)$, (d) 分離抽出された信号 $\hat{f}_1(t)$

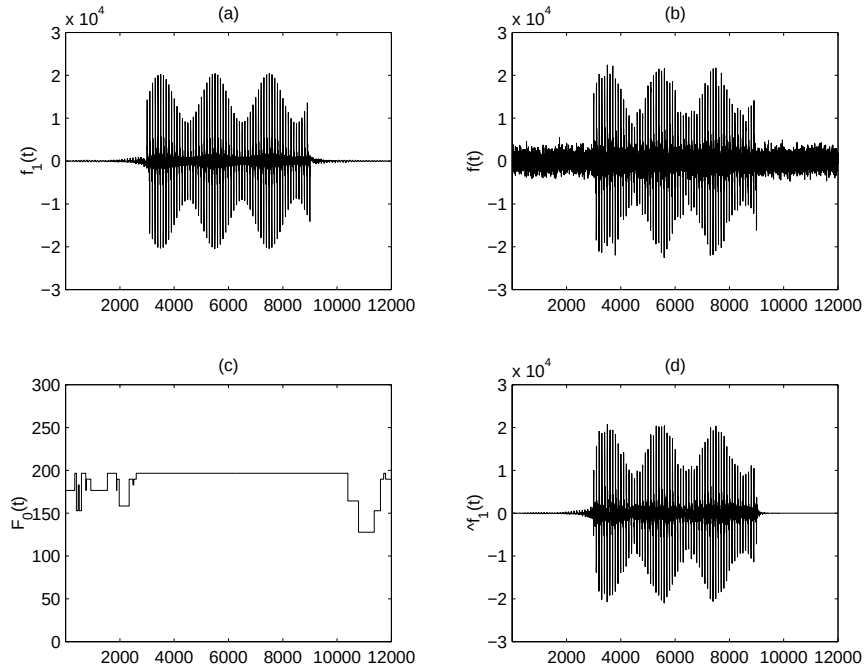

図 4.16: 分離例（Condition 1, AM 調波複合音＋ランダム帯域雑音）：(a) 原信号/a/ $f_1(t)$, (b) 混合信号 $f(t)$, SNR= 10 dB, (c) 推定された基本周波数 $F_0(t)$, (d) 分離抽出された信号 $\hat{f}_1(t)$

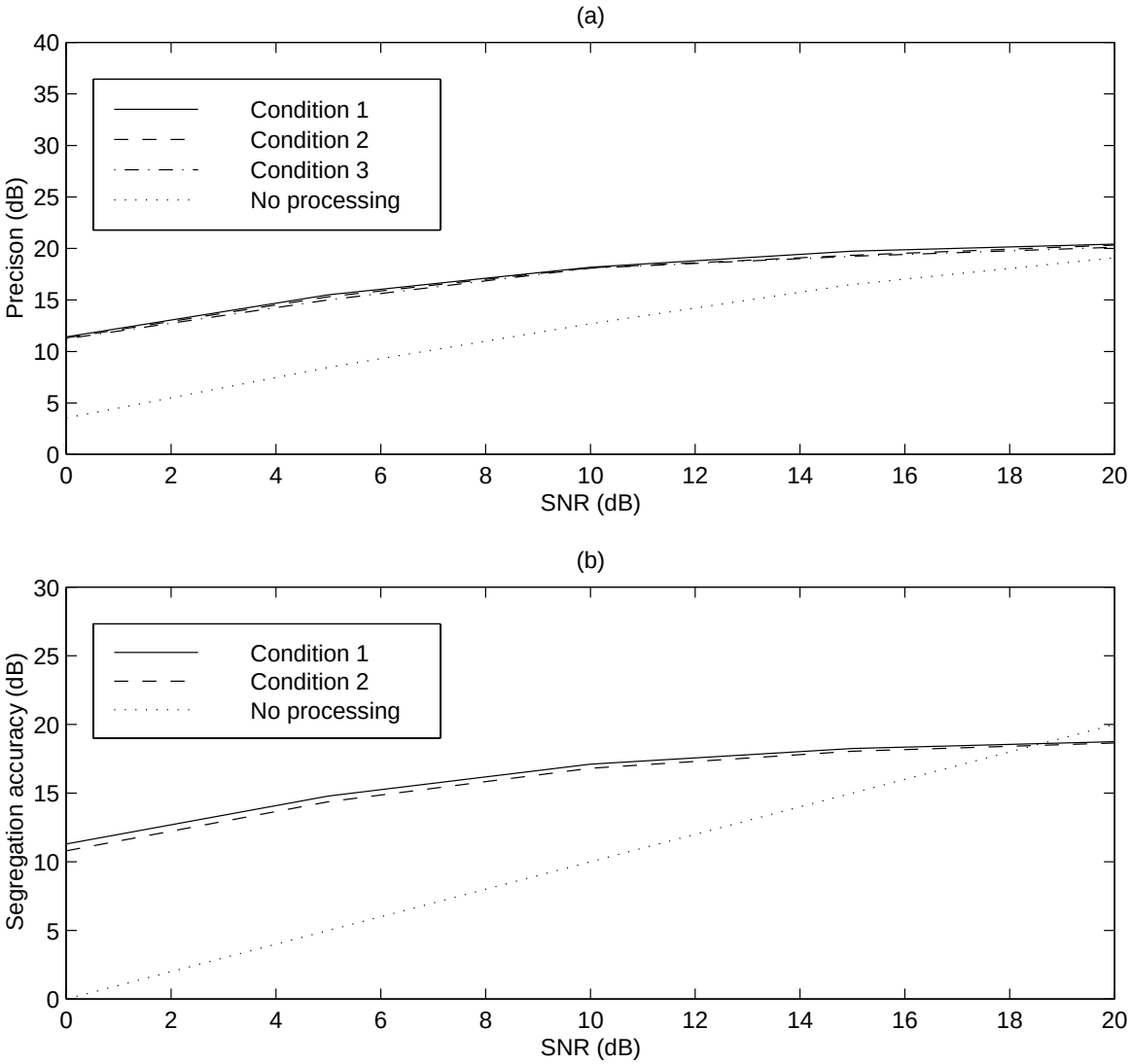

図 4.17: 分離精度の比較（合成母音＋ピンク帯域雑音の場合）: (a) Precision, (b) Segregation accuracy

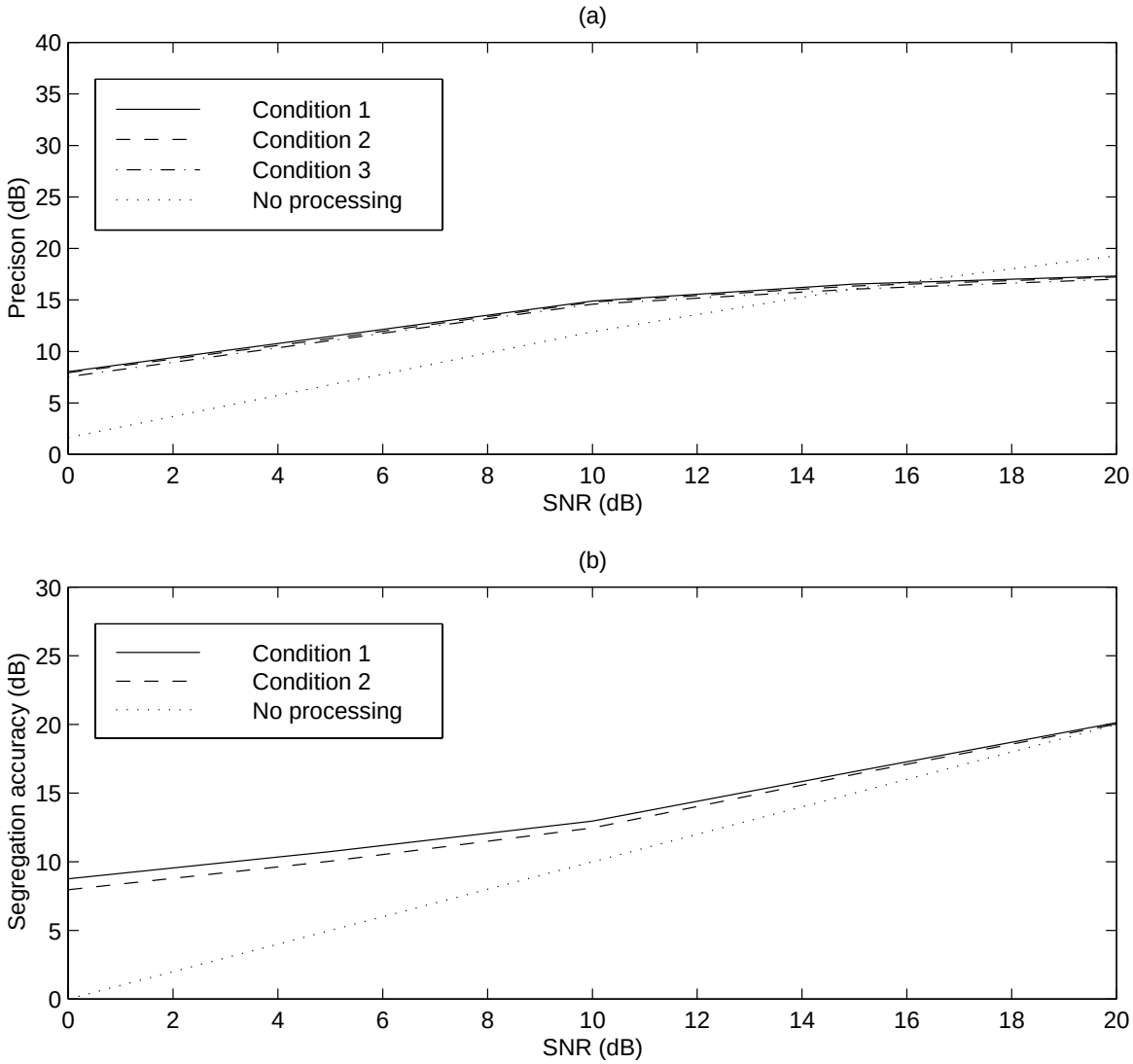

図 4.18: 分離精度の比較（合成母音＋ランダム帯域雑音の場合）：(a) Precision, (b) Segregation acuuracy

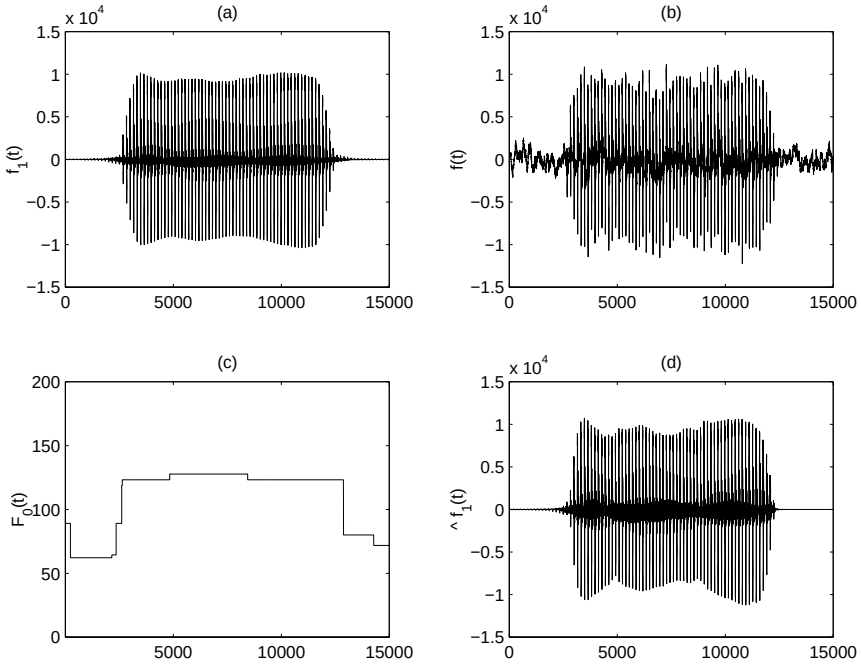

図 4.19: 分離例（Condition 1, 合成母音＋ピンク帯域雑音の場合）: (a) 原信号/a/ $f_1(t)$, (b) 混合信号 $f(t)$, SNR= 10 dB, (c) 推定された基本周波数 $F_0(t)$, (d) 分離抽出された信号$\hat{f}_1(t)$

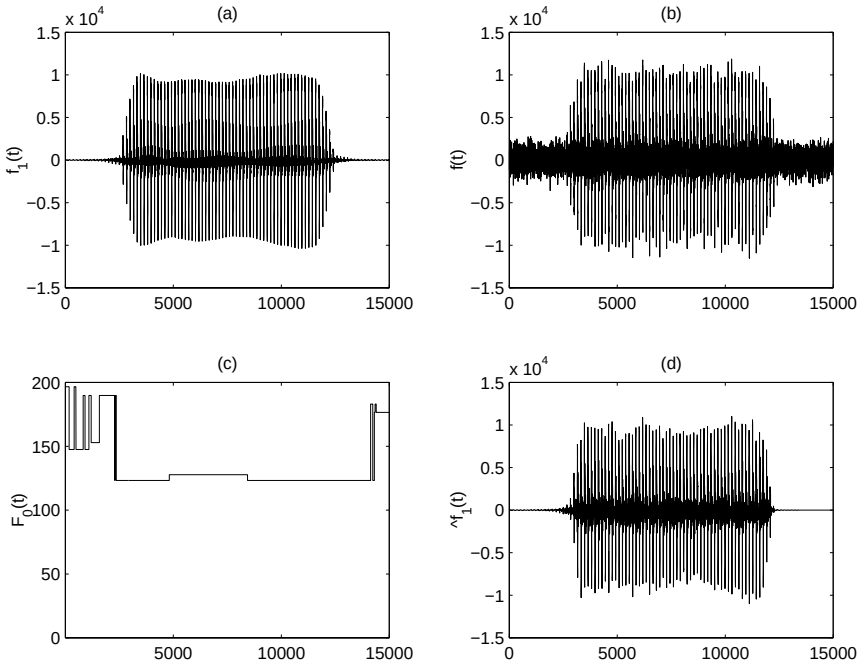

図 4.20: 分離例（Conditon 1, 合成母音＋ランダム帯域雑音の場合）: (a) 原信号/a/ $f_1(t)$, (b) 混合信号 $f(t)$, SNR= 10 dB, (c) 推定された基本周波数 $F_0(t)$, (d) 分離抽出された信号 $\hat{f}_1(t)$

向上が見られる。例えば、図 4.19 (a) の合成母音に SNR= 10 dB のピンク帯域雑音が付加されたとき、混合信号は図 4.19 (b) となる。次に、分析フィルタ群の出力から基本周波数が図 4.19 (c) に示すように推定され、第３章で提案された解法により合成母音が図 4.19 (d) に示すように分離抽出される。

最後に、合成母音とランダム帯域雑音が付加された混合信号の分離結果を図 4.18に示す。妨害音がランダム帯域雑音の場合でも、SNR が良好な場合（20 dB）は若干の分離精度の向上しか見られないが、SNR が最悪な場合（0 dB）は顕著に分離精度の向上が見られる。ただし、若干、ランダム帯域雑音の場合に全体の分離精度が低下している。例えば、図 4.20 (a) の合成母音に SNR= 10 dB のピンク帯域雑音が付加されたとき、混合信号は図 4.20 (b) となり、分析フィルタ群の出力から基本周波数が図 4.20 (c) に示すように推定され、本章で考案した解法により合成母音が図 4.20 (d) に示すように分離抽出される。

以上の結果から、$D_{k,0}$の定数制約では、波形レベルで完全に復元できないため、Segregation accuracy での評価はできなかった。興味あることに、Precision の評価尺度では瞬時位相の時間変動に対する区分多項式近似の次数による影響はほとんどなかった。この考察から、振幅スペクトルのみを十分に分離抽出するだけでよければ、瞬時位相の時間変動の区分多項式近似の次数を落してもさほど振幅スペクトルの分離精度には大きな影響を与えないといえる。しかし、当然のことながら、波形レベルで十分な精度を持つように復元するためには、瞬時位相の時間変化を区分１次多項式で制約する必要がある。

4.4.6　考察

図 4.15〜図 4.18の結果を比較すると、

- 調波複合音が周波数変調の有無に関わらず、五つの制約条件を利用することで十分に分離抽出が可能である。
- 妨害雑音として二種類の雑音を用意したが、ランダム帯域雑音の場合に分離精度が若干、低下した。

また、図 4.14における振幅スペクトルの分離の程度を調べてみる。この比較結果を図 4.21 に示す。ここで、図中の振幅スペクトルは、フレーム長 51.2 msec、hamming 窓とした短時間 Fourier 変換により求めたものである。図 4.21 (a) は原信号と分離抽出後の振幅スペクトルの比較を、図 4.21 (b) は原信号と混合信号との振幅スペクトルの比較を示している。この図から、雑音成分がどれだけ正確に分離抽出できているか、言い換えると雑音を除去できているか視覚的によく理解できる。

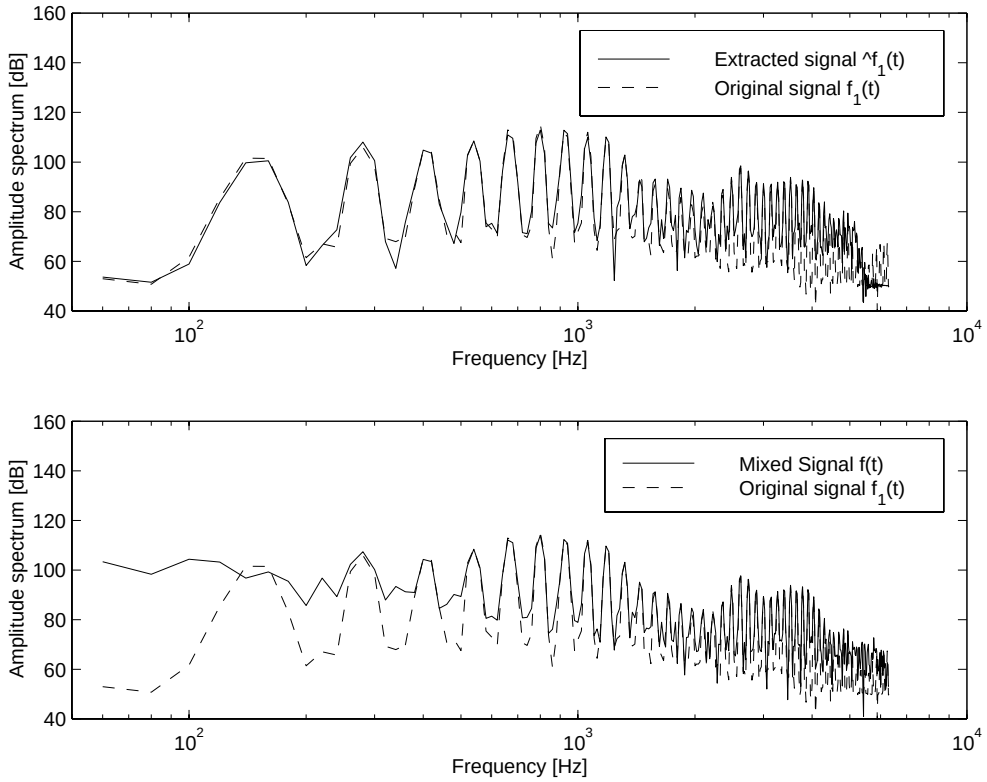

図 4.21: 振幅スペクトルの比較：(a) 原信号 $f_1(t)$ と分離抽出した信号$\hat{f}_1(t)$ の比較, (b) 原信号 $f_1(t)$ と混合信号 $f(t)$（SNR=10 dB のピンク雑音）の比較

以上の考察により、本章で考案した二波形分離問題の解法は、周波数変調の有無に関係なく、目的の調波複合音を分離抽出でき、十分性を満たしていることがわかる。また、第３章で提案した二波形分離問題の解法を利用することで、AM–FM 調波複合音をより複雑にした音声信号の分離抽出も可能になるものと考えられる。

4.5 制約条件の有効性の検証

上記 4.3 節および 4.4 節の検証から、二波形分離問題で利用した制約条件の十分性を示した。次に、残る項目である制約条件の有効性の検証を行う。

4.5.1 検証シミュレーションの条件

発展的構成法に従い、本論文で採用した制約条件を一つずつ省略した場合の分離精度を検証することで、制約条件の有効性を検証する。

そこで、本節では、AM–FM 調波複合音の分離抽出における二波形分離問題を議論しているため、次の四種類の混合信号に対する二波形分離問題：

1. AM 調波複合音＋ピンク帯域雑音
2. AM 調波複合音＋ランダム帯域雑音
3. LMA 合成母音＋ピンク帯域雑音
4. LMA 合成母音＋ランダム帯域雑音

を考える。次に、四つの検証条件：

- 提案方法：すべての制約条件の利用
- Condition 1: Comb filter による調波成分抽出＋ Kalman filter で求めた $C_{k,0}(t)$ と $D_{k,0}(t)$ だけの利用
- Condition 2: Comb filter による調波成分抽出
- Condition 3: 処理なし（分析合成系による全域通過）

の比較を含め、二波形分離シミュレーションにより分離精度を求めることで、制約条件の有効性を検証する。ここで、Condition 1 は、制約条件 (2.2) のなめらかさを省略した場合、Condtion 2 は制約条件 (2.1) の区分多項式近似と制約条件 (2.2) のなめらかさを省略した場合、Condition 3 は、すべての制約条件を省略したものである。

4.5.2 検証結果

はじめに、AM 調波複合音とピンク帯域雑音を混合した中から AM 調波複合音を分離抽出する問題を考える。雑音の SNR を 0〜20 dB まで 5 dB 刻に変化させ、四つの検証条件と二つの評価尺度で比較した結果を図 4.22に示す。ここで、Precision の評価量と Segregation accuracy の評価量は高ければ高いほど精度が良いことを示す。このことから、特性図における四つの検証条件の結果は、二つの図で左上がりのグラフを描けば、本論文で利用した制約条件の有効性を示すことができる。

本解法と三つの検証条件（Condition 1、2、3）の比較を行ったところ、図 4.22の結果から、いずれも本解法の分離精度が一番高いことがわかる。

ここで、Condition 1 と本解法の精度を比較すると、なめらかさ（制約条件 (2.2)）の制約を利用したことによる分離精度の向上を確認できる。また、本解法と Condition 1、およ

び Condition 2 の比較では、同一周波数領域に二波形の成分が存在する際、位相情報を利用したことによる分離精度の向上を確認できる。特に、本解法と Condition 3 の比較では、各評価尺度の改善量を求めることで、本解法の雑音除去性能を求めることができる。

他、同様に、AM 調波複合音とランダム帯域雑音を混合した場合、LMA 合成母音とピンク帯域雑音を混合した場合、LMA 合成母音とランダム帯域雑音を混合した場合について、それぞれ、図 4.23、図 4.24、図 4.25に示す。

以上、すべての検証結果について、本解法と三つの検証条件（Condition 1、2、3）の比較を行ったところ、いずれも本解法の分離精度が一番高いことがわかる。

この結果、不良設定の逆問題である二波形分離問題を一意に解くためには、分離抽出したい信号の瞬時振幅、瞬時位相、基本周波数の時間変化に着目し、表 5.1に上げた五つの数理工学的な制約条件を用いて解けばよいといえる。

4.6 むすび

本章では、第３章で提案した二波形分離問題の解法で利用した制約条件の十分性および有効性を検証した。

はじめに、分離抽出の対象となる音を、AM 単一成分音とした場合について、二波形分離問題の解法による分離精度を評価した。ここでは、単一成分音の瞬時振幅 $A_k(t)$ に対し、漸近的変化の多項式近似となめらかさの制約条件を検証した。主に、瞬時振幅の時間変動に対する区分多項式近似の表現精度について分離精度を評価した。この結果、振幅包絡 $A_k(t)$ を微小区間で定数と見なし、漸近的変化のなめらかさを境界条件としてみた場合、純音の分離抽出は十分性を満たしたものの、振幅包絡が上下に変動するような AM 単一音では十分性の意味で分離抽出できないことがわかった。この結果、第３章で定式化した制約条件の有効性を示すことができた。

以上の結果から、振幅包絡を正確に分離抽出するための十分性は、振幅包絡の時間変動の制約条件を１次の区分的多項式で近似することであることがわかった。

次に、分離抽出の対象となる音を AM–FM 調波複合音とした場合について、二波形分離問題の解法による分離精度を評価した。特に基本周波数が時間的に変動するものとそうでないものの二種類を用意し、瞬時位相の時間変動に対する区分多項式近似の表現精度について分離精度を評価した。また、同時に、複合成分における瞬時振幅と基本周波数の時間変動の有無による影響も検証した。この結果、位相の時間変動の拘束は、波形レベルで十分な精度をもつようにに復元するためには、１次の区分多項式近似で拘束することが十分条件となるが、振幅スペクトルのレベルでの分離であれば、定数項による近似あるいは 0

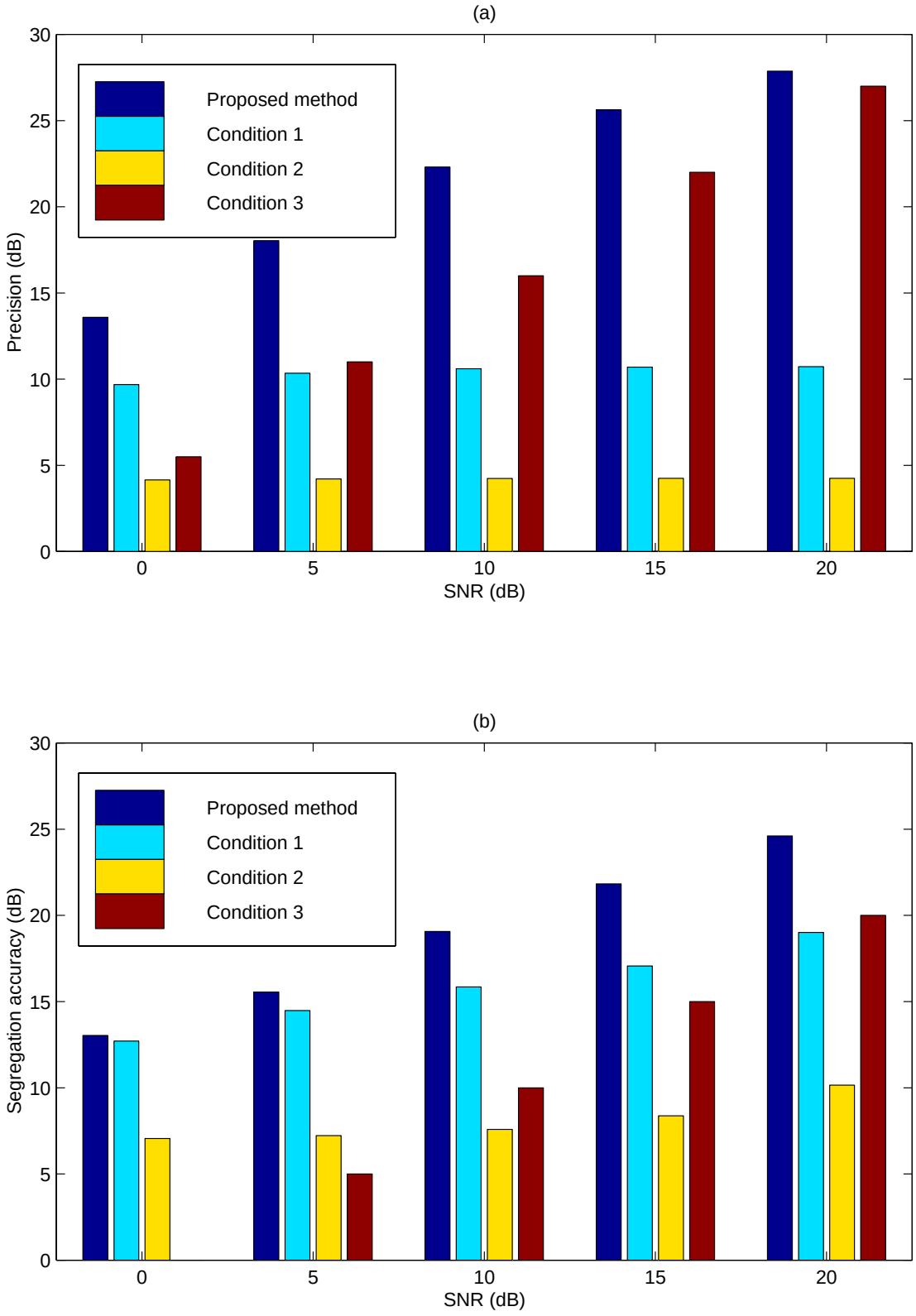

図 4.22: 分離精度の比較（AM 調波複合音とピンク帯域雑音の混合）：(a) Precision, (b) Segregation accuracy

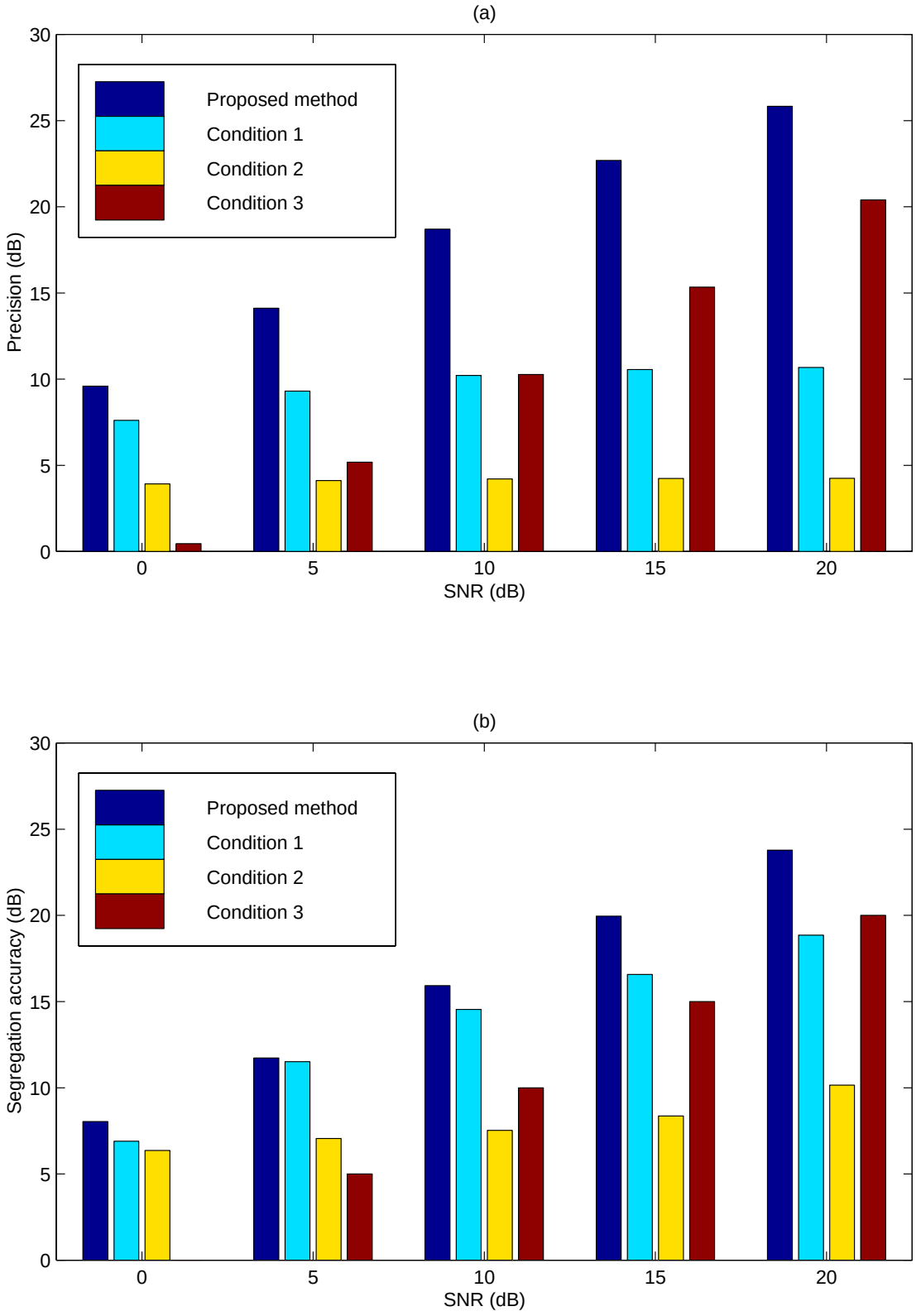

図 4.23: 分離精度の比較（AM 調波複合音とランダム帯域雑音の混合）：(a) Precision, (b) Segregation accuracy

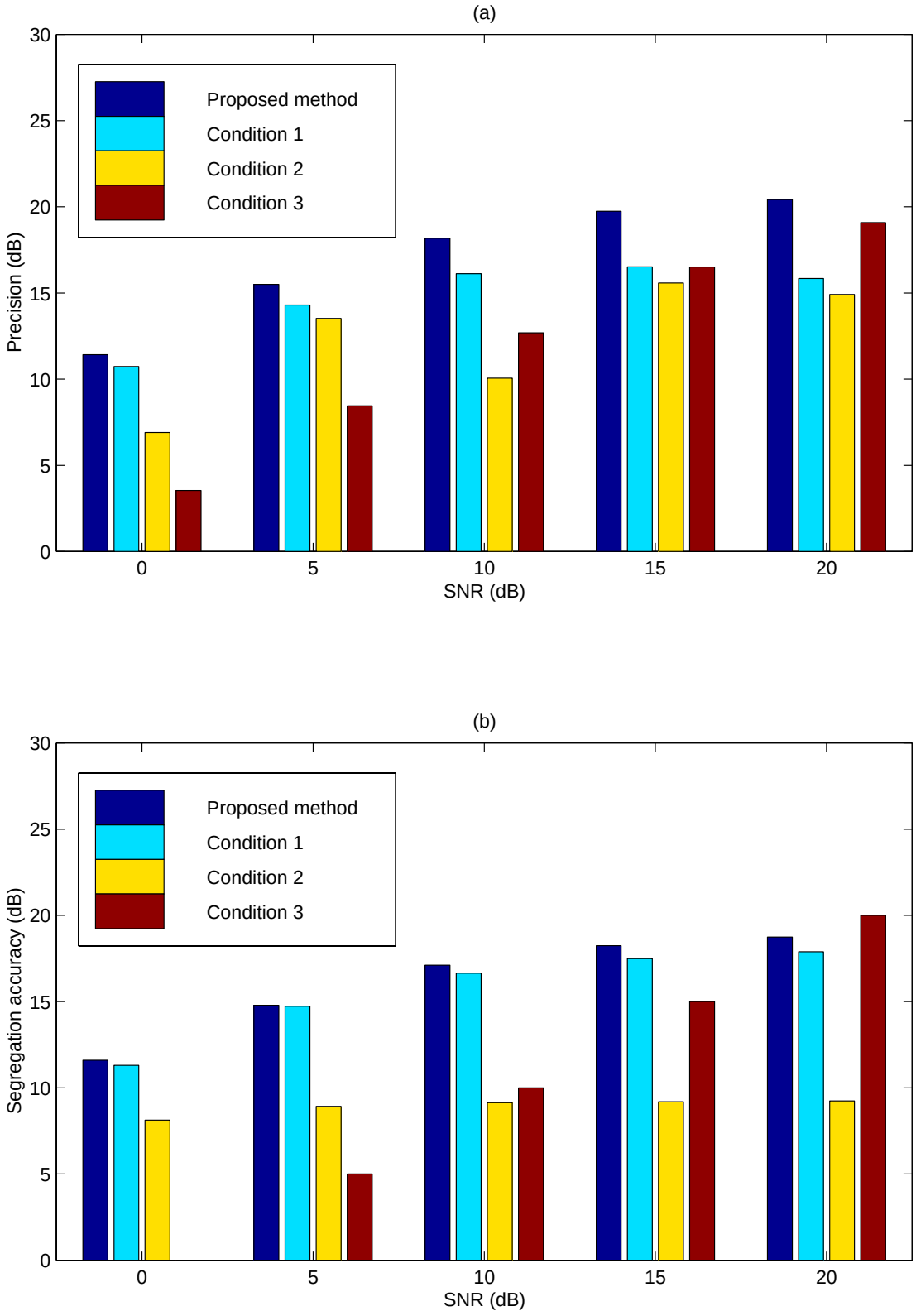

図 4.24: 分離精度の比較（LAM 合成母音とピンク帯域雑音の混合）：(a) Precision, (b) Segregation accuracy

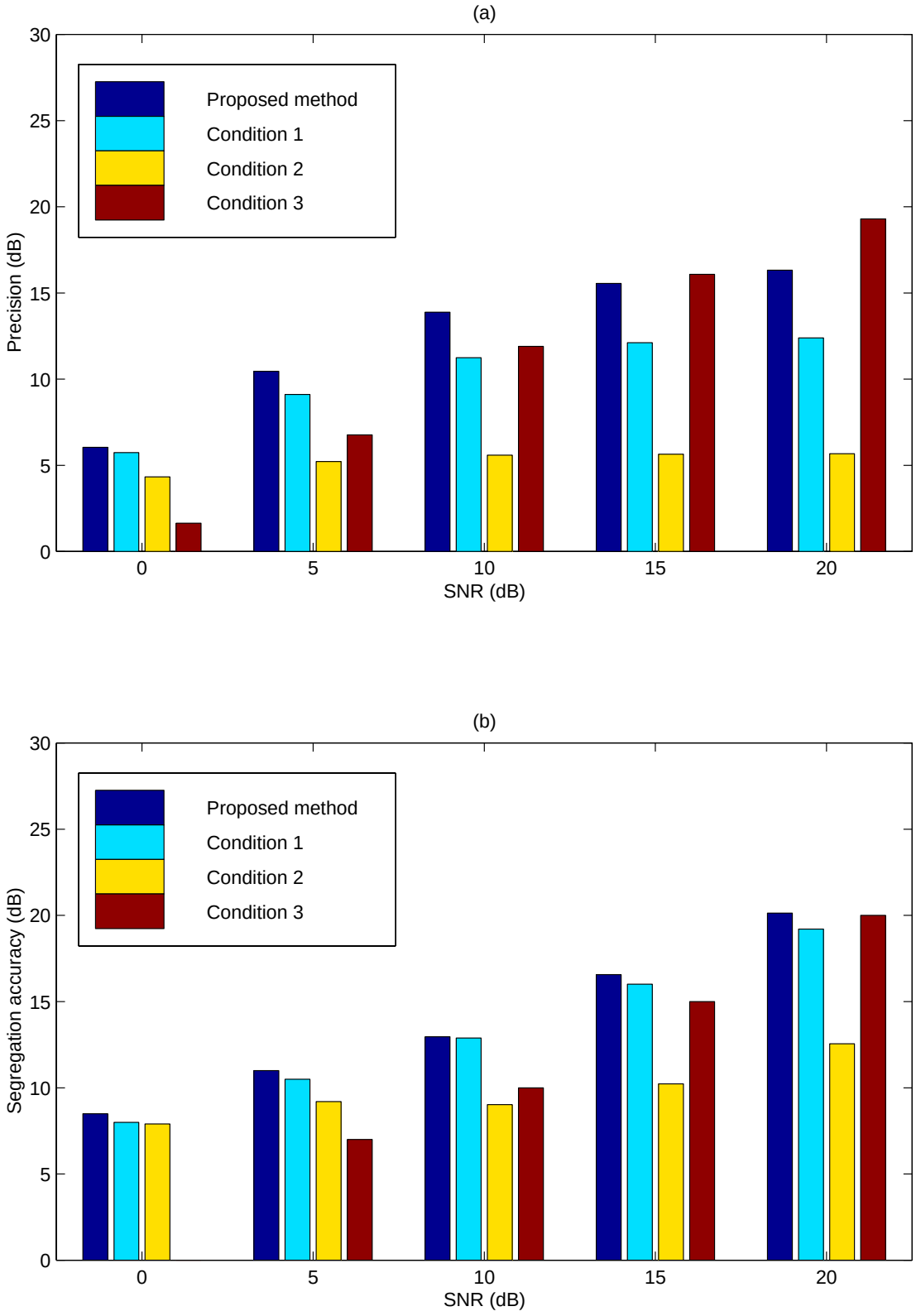

図 4.25: 分離精度の比較（LAM 合成母音とランダム帯域雑音の混合）：(a) Precision, (b) Segregation accuracy

次の区分多項式の近似が十分条件となった。また、分離抽出の対象となる音が周波数変調されているかどうかにより、分離精度に影響があるかどうかを調べたところ、ほとんど影響はみられなかった。つまり、基本周波数の時間変動の拘束が区分的に一定であるという制約で十分であることがわかった。

以上をまとめると、対象音を AM–FM 調波複合音に拡張した場合、瞬時振幅と瞬時入力位相の時間変動に対する区分多項式近似の拘束、瞬時振幅と瞬時入力位相のなめらかさの拘束、調波関係と共通の立上り・立下りの他、基本周波数の時間変動に対する区分多項式近似の拘束と、第３章で提案したすべての制約条件があれば、AM–FM 調波複合音を分離抽出できることがわかった。このときに利用したすべての制約条件は二波形分離問題を解くための十分条件であった。

最後に、AM–FM 調波複合音を利用し、二波形分離問題の解法で利用した制約条件を順次省略したときの分離精度を評価することで、制約条件の有効性を検証した。この結果、Precision および Segregation accuracy の評価尺度すべてにおいて、すべての制約条件を利用すること（第３章で提案した解法）の有効性が示された。

第 5 章

音の分離抽出における聴覚の計算の方略の提案

5.1 まえがき

本章では、第４章で検証された二波形分離問題の解法から、発展的構成法に従って聴覚の計算の方略の構築を試みることを目的とする。

はじめに、第３章で提案し、第４章でその十分性と有効性が検証された二波形分離問題の解法と制約条件、およびアルゴリズムの実装について総括する。ここでは、分離抽出したい信号を AM–FM 調波複合音とおき、雑音中からこの調波複合音を分離抽出する二波形分離問題の解法を説明する。

最後に、この解法から、音の分離抽出における聴覚の計算の方略を提案する。

5.2 二波形分離問題の解法の総括

第３章で提案された二波形分離問題の解法は、第４章で、発展的構築法に基づき、その十分性と有効性について検証された。第４章前半では、第３章で提案した二波形分離問題の解法が、AM–FM 調波複合音の二波形分離問題を一意に解く方法として十分性を満たすことを示した。また、第４章後半では、二波形分離問題の解法で利用する制約条件の有効性を示した。以上の結果から、第３章で提案した二波形分離問題の解法は、「どのような制約条件を用いることで二波形分離問題を一意に解くことができるか」という戦略的な解法を示している。

そこで、次に、この解法で利用した音の分離抽出に必要な物理量と制約条件について概要をまとめる。

5.2.1 二波形分離問題における仮定と制約条件

本章では、$f_1(t)$ を AM–FM 調波複合音と仮定し、雑音 $f_2(t)$ 中に $f_1(t)$ が加算される状態から $f_1(t)$ を分離抽出する問題とする。また、この調波複合音は、基本周波数 $F_0(t)$ を整数倍した高調波成分をもつものとする。

次に、二波形分離問題で利用した制約条件と Bregman によって提唱された四つの発見的規則の対応関係を表 5.1に示す。Bregman によって提唱された四つの発見的規則 [Bregman, 1993] は、表 5.1の左側に示されるように、我々の経験する環境に存在する音とはどういうものなのかを述べているのに等しい。しかし、これらは定性的なものであるため、制約条件式として直接利用することができなかった。そこで、分離抽出の対象となる音をより複雑なものに拡張しつつ、それに応じて二波形分離問題を一意に解くために必要な四つの発見的規則を、表 5.1の右側に示す対応関係で、数理工学的な制約条件にとらえ直した。ま

表 5.1: Bregman 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	制約条件式
(i) 関連の無い音が一緒に始まったり、終ったりすることはない	(1) 立上り・立下りの同期	$\lvert T_{\mathrm{S}} - T_{k,\mathrm{on}}\rvert \leq \Delta T_{\mathrm{S}}$ $\lvert T_{\mathrm{E}} - T_{k,\mathrm{off}}\rvert \leq \Delta T_{\mathrm{E}}$
(ii) 変化は急激には起こらない	(2) 漸近的変化	
(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	(2.1) 区分多項式近似（ゆっくりと）	$dA_k(t)/dt = C_{k,R}(t)$ $d\theta_{1k}(t)/dt = D_{k,R}(t)$ $dF_0(t)/dt = E_{0,R}(t)$
(b) 同じ音源から生じる音の一連の音の属性は、ゆっくりとなめらかに変化する傾向にある	(2.2) なめらかさ	$\sigma_A = \int_{t_a}^{t_b} [A_k^{(R+1)}(t)]^2 dt \Rightarrow \min$ $\sigma_\theta = \int_{t_a}^{t_b} [\theta_{1k}^{(R+1)}(t)]^2 dt \Rightarrow \min$
(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的成分が発生する	(3) 調波関係	$n \times F_0(t), \quad n = 1, 2, \cdots, N_{F_0}$
(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	(4) 振幅包絡 $A_k(t)$ 間の相関	$\frac{A_k(t)}{\lVert A_k(t)\rVert} \approx \frac{A_\ell(t)}{\lVert A_\ell(t)\rVert}, \qquad k \neq \ell$

ず、制約条件 (1) は各分析フィルタ出力で得られた立上り・立下りと基本波の立上り・立下りの一致の誤差を拘束するものである。次に制約条件 (2) は、分離抽出したい信号の瞬時振幅と瞬時入力位相、および基本周波数の時間変化を拘束するものである。特に、制約条件 (2.1) では時間変化を区分多項式で近似し、制約条件 (2.2) では近似による時間変化のなめらかさを拘束する。制約条件 (3) は、基本周波数の倍音関係にある成分を拘束し、制約条件 (4) は各分析フィルタ出力における振幅包絡間の相関を拘束するものである。

これらは、第３章で定式化されたが、第４章で、二波形分離問題の解法と制約条件を検証した結果、表 5.1の制約条件が二波形分離問題を一意に解くための十分条件であり、かつ有効であることが明らかになった。

5.2.2 解法アルゴリズムの概要

本論文で提案する二波形分離問題の解法は、図 3.3に示すモデルで実現され、図 5.1に示す順序で処理が行われる。ここで、図 3.2に示す二波形分離モデルは、最終的に、(a) 分析フィルタ群、(b) 基本周波数の推定部、(c) 波形分離部、(d) グルーピング部の４ブロックで構成された。また、二波形分離問題を解くための一連の流れは次のとおりである。

はじめに、混合信号 $f(t)$ のみが観測され（図 5.1. A）、分析フィルタ群により、瞬時振

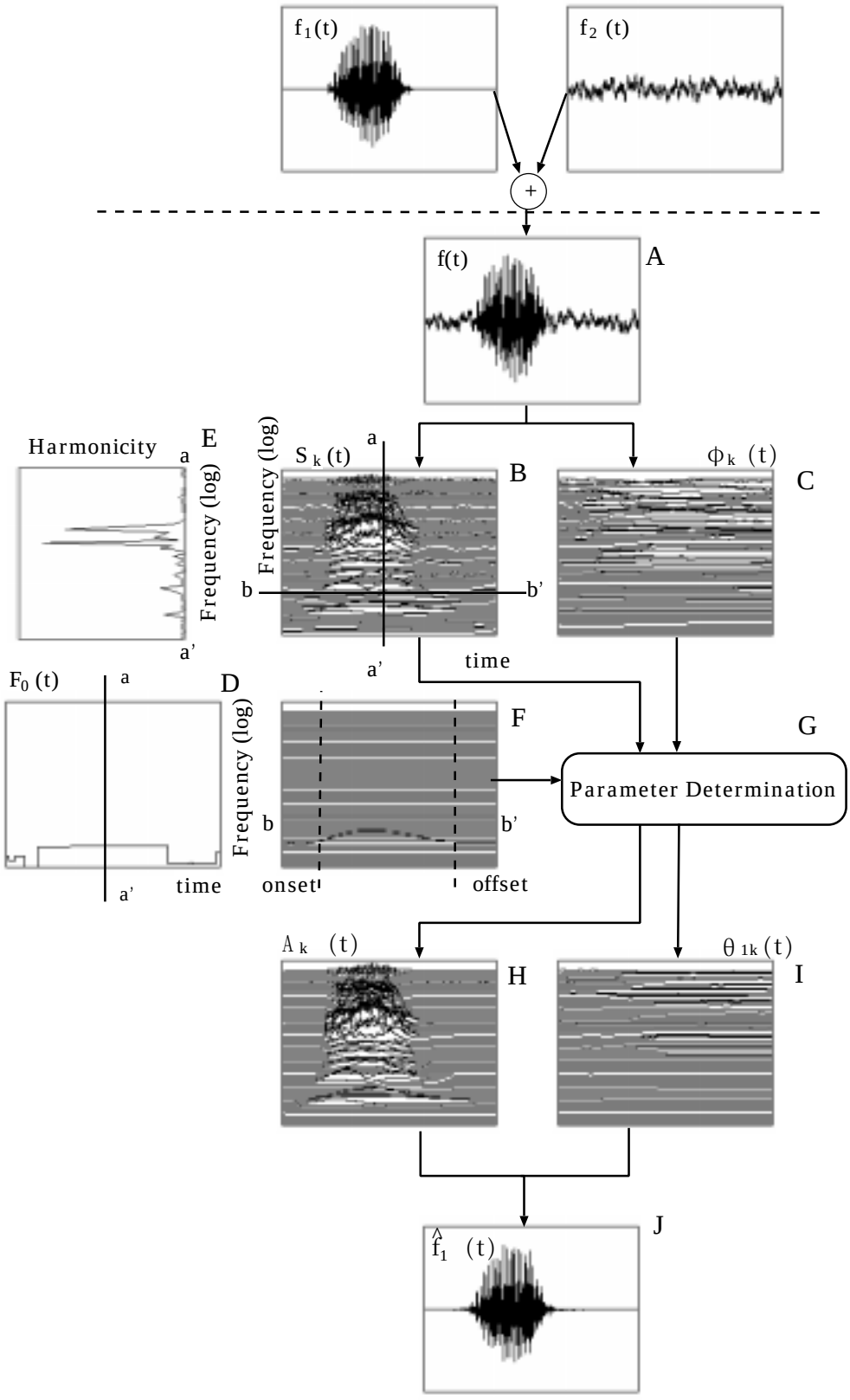

図 5.1: 二波形分離モデルの信号処理の概要

幅 $S_k(t)$ と瞬時出力位相 $\phi_k(t)$ に分解される（図 5.1. B、C）。次に、$S_k(t)$ から基本周波数 $F_0(t)$ を求め（図 5.1. D）、二波形分離の対象となる時間–周波数領域を決定する（3.2.2 節参照）。調波成分の存在する周波数領域については、$F_0(t)$ と発見的規則 (iii) の調波関係（図 5.1. E、a–a'）を用いて決定する。調波成分の存在する時間領域については、発見的規則 (i) の、各高調波成分の立上りと立下りの同期（図 5.1. F、b–b'）を用いて決定する。

次に、波形分離部では、上記で決定された時間–周波数領域において $S_k(t)$ と $\phi_k(t)$ から望みの信号の $A_k(t)$ と $\theta_{1k}(t)$ を求める（図 5.1. G）。これは、$A_k(t)$ と $\theta_{1k}(t)$ を発見的規則 (ii) の漸近的変化（ゆっくりと）を用いて最適化問題として解く（図 5.1. H、I）。但し、最適解の候補が多過ぎるため、発見的規則 (ii) の漸近的変化（なめらかさ）を加えて採用し解の探索範囲を狭め、発見的規則 (iv) の振幅包絡間の変動の一致（相関）を手がかりとして最適解の絞り込みを行う。

最後に、グルーピング部では、$A_k(t)$ と $\theta_{1k}(t)$ がグルーピングされ、合成フィルタ群を用いて $\hat{f}_1(t)$ に再構成される（図 5.1. J）。図中では割愛しているが、$\hat{f}_1(t)$ と同様に、$B_k(t)$ と $\theta_{2k}(t)$ がグルーピングされ、合成フィルタ群を用いて $\hat{f}_2(t)$ に再構成される。

以上が二波形分離問題を一意に解くためのアルゴリズムの概要である。

5.3 音の分離抽出における聴覚の計算の方略

雑音下からの音の分離抽出という二波形分離問題を解くために、発展的構成法に従い、制約条件とモデルの発展的な検証を行ってきた。前節の検証実験により、二波形分離問題を一意に解くために有効な制約条件として、四つの発見的規則を数理工学的な制約条件にとらえ直し、これを利用することで、聴覚の情景解析で説明できる音の分離抽出の計算モデルを実現することができた。この結果、不良設定の逆問題である二波形分離問題を一意に解くためには、分離抽出したい信号の瞬時振幅、瞬時位相、基本周波数の時間変化に着目し、表 5.1に上げた五つの数理工学的な制約条件を用いて解けばよいといえる。

ここで、第２章で述べた四つの発見的規則の思想と音とは何であるかということを再考してみる（図 5.2参照）。発見的規則が生態学的アプローチ以外の何者でもないことから、我々は身の回りの環境から、四つの発見的規則に従う音のかたまりを一つの音源で生じた音と知覚するわけである。さて、我々の身の回りの環境が本論文で定義した二波形分離問題と想定すれば、本論文で有効性が示された制約条件はどんな意味になるであろうか。答えは明白である。二波形分離問題を一意に解くために利用した制約条件の意味は、分離抽出したい音を四つの発見的規則に該当する音として取り出すことである。もう少し詳細に述べると、制約条件の意味は、分離抽出したい音の物理量に対し、時間的変動を漸近的変

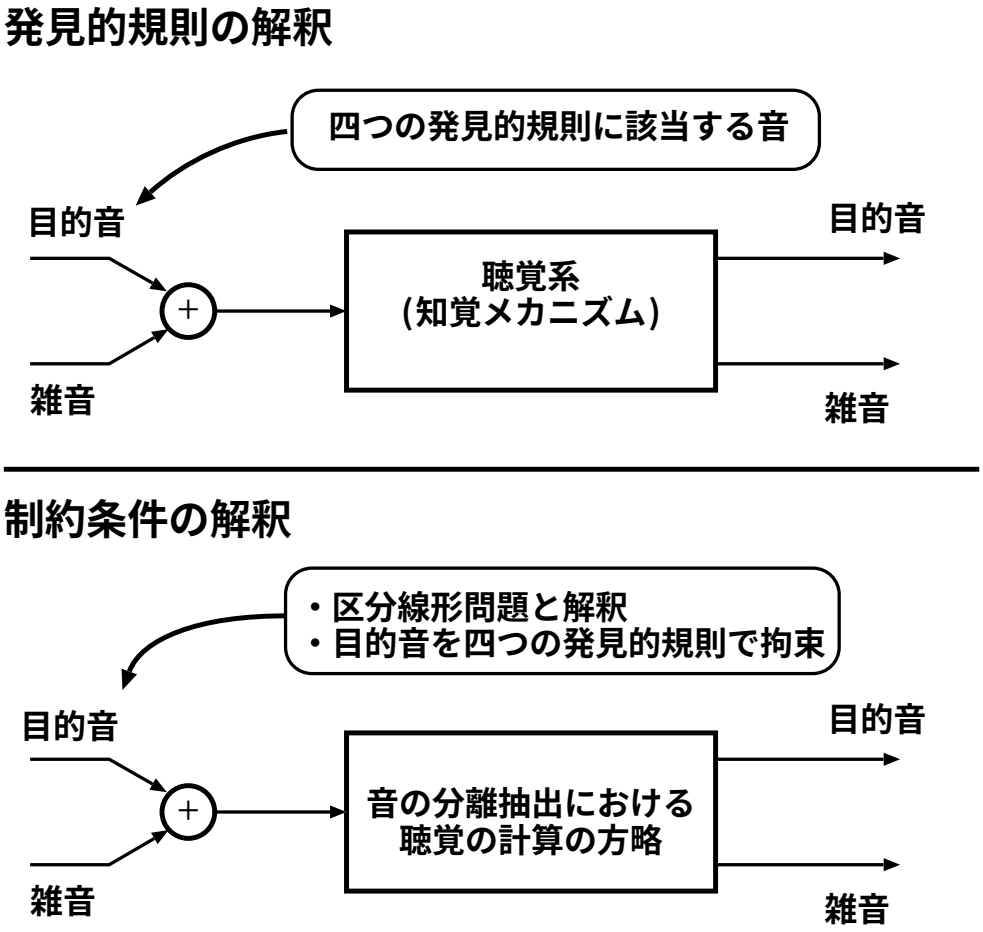

図 5.2: 制約条件の解釈

化の発見的規則で拘束し、その拘束された物理量から更に残りの発見的規則を利用して一意な音の物理量を決定することである。以上は、聴覚の情景解析に基づいた制約条件の意味の解釈である。

次に、二波形分離問題の制約条件を数理工学的な意味で解釈してみる（図 5.2）。二波形分離問題で利用した制約条件の意味は、この問題を区分線形問題と見なし、分離抽出したい音の各物理量の時間変動を拘束することで目的の音を分離抽出することである。更に、詳細に述べると、制約条件を利用する意味は、不良設定問題である二波形分離問題を区分線形問題と見なし、瞬時振幅、瞬時位相、基本周波数の時間変化を区分多項式近似で拘束し、それぞれがなめらかであるものから、振幅包絡間の相関が最大になるように瞬時振幅、瞬時位相、基本周波数を取り出すことである。以上は、数理工学的立場からの制約条件の解釈であり、これは言い換えると、二波形分離問の解法を通じて、制約条件を十分条件とした、音の分離抽出に関する聴覚の処理機能の「入力、出力、処理過程」を示したことになる。この関係を表 5.2に示す。ここで、入力は、分離抽出したい信号の物理量（瞬時振

表 5.2: 音の分離抽出における計算の方略

入力	分離抽出したい信号の物理量（瞬時振幅、瞬時位相、基本周波数）
出力	信号波形（あるいは瞬時振幅、瞬時位相、基本周波数の物理量）
処理過程	不良設定問題を区分線形問題と見なし、各物理量の時間変化を拘束し、それぞれがなめらかであるものから振幅包絡間の相関が最大になるように各物理量を取り出すこと。

幅、瞬時位相、基本周波数）であり、出力は信号波形である。出力を信号波形とする理由は、波形レベルで復元できるほど正確に分離を行うことを狙いとしているためである。しかし、聴覚系が信号を分離抽出した後、波形レベルに復元しているとは考え難いため、本論文では、分離抽出した信号を波形レベルに復元可能な物理量も出力と見なす。

表5.2に示した処理機能は、不良設定問題である二波形分離問題を一意に解くということを両者の立場から統一的に議論した結果であり、「どのような制約条件を用いることで二波形分離問題を一意に解くことができるか」という戦略的な解法を示している。また、二波形分離問題の解法で利用した制約条件は、心理学的に意味のある制約条件であり、その十分性と有効性も議論されている。従って、本論文で提案した二波形分離問題の解法は、音の分離抽出における聴覚の計算の方略を示したことになる。

以上をまとめると、本論文で提案した音の分離抽出における聴覚の計算の方略とは、音の分離抽出という不良設定問題を区分線形問題と見なし、分離抽出したい信号の物理量の時間変化、つまり音の物理量の動きを拘束することで一意に解く、ということである。本論文では、計算の方略を導く際、制約条件の十分性と有効性しか議論していないが、制約条件の必要十分性を示すことで計算の方略を計算理論に発展させることができる。この必要十分性を導くためには、沢山の計算の方略を提案し、聴覚心理実験・生理実験によりこれらを検証することで正しいものに絞り込まなければならない。そのため、検証に多くの時間を必要とするが、本論文では、聴覚の計算理論を構築するための明確な方法論を提供できたと同時に、それに向けて確実に一歩前進したといえる。

5.4 むすび

本章では、第４章で検証された二波形分離問題の解法から、発展的構成法に従って聴覚の計算の方略の構築を試みた。

はじめに、第３章で提案され、第４章でその十分性と有効性が検証された二波形分離問題の解法と制約条件、およびアルゴリズムの実装について総括した。特に、分離抽出したい信号を AM–FM 調波複合音とし、雑音中からこの調波複合音を分離抽出する二波形分離問題を説明した。このとき、本論文で得られた解法は、分離抽出したい信号の瞬時振幅、瞬時入力位相、基本周波数の時間変動を拘束し、Bregman によって提唱された四つの発見的規則に対応した数理工学的な制約条件を用いることであった。

本章で総括した解法は、「不良設定問題である二波形分離問題において、制約条件を用いてどのように積極的に一意な解を求めようとするか」という戦略的な意味が込められた解法である。この事実から、本解法は音の分離抽出における聴覚の計算の方略を示した。

まず、聴覚の情景解析の立場から制約条件の意味を述べると、この解法は、分離抽出したい音の物理量に対し、時間的変動を漸近的変化の発見的規則で拘束し、その拘束された物理量から更に残りの発見的規則を利用して一意な音の物理量を決定することであった。また、数理工学的立場から制約条件の意味を述べると、これは、不良設定問題である二波形分離問題を区分線形問題と見なし、瞬時振幅、瞬時位相、基本周波数の時間変化を区分多項式近似で拘束し、それぞれがなめらかであるものから、振幅包絡間の相関が最大になるように瞬時振幅、瞬時位相、基本周波数を取り出すことであった。従って、検証された二波形分離問題の解法は、不良設定問題を一意に解くという問題を両者の立場から統一的に議論して得られたものであるため、本解法を聴覚の計算の方略と解釈できた。

以上の結果、本論文では、音の分離抽出における聴覚の計算の方略を、音の分離抽出という不良設定問題を区分線形問題と見なし、分離抽出したい信号の物理量の時間変化、つまり動きを拘束することで一意に解くことである、と結論づけた。本論文では、計算の方略を導く際、制約条件の十分性と有効性しか議論していないが、制約条件の必要十分性を示すことで計算の方略を計算理論に発展させることができる。この必要十分性を導くためには、沢山の計算の方略を提案し、聴覚心理実験・生理実験によりこれらを検証することで正しいものに絞り込まなければならない。そのため、検証に多くの時間を必要とするが、本論文では、聴覚の計算理論を構築するための明確な方法論を提供できたと同時に、それに向けて確実に一歩前進したといえる。

第 6 章

音の分離抽出における聴覚の計算の方略の正当性

6.1 まえがき

本章では、(1) 実音声（母音）を対象とした二波形分離問題、(2) 共変調マスキング解除を想定した二波形分離問題、という実際的な二波形分離問題に対し、本論文で提案した計算の方略を展開することで、本計算の方略がこれらの問題の解法を導出できることを示す。

6.2 実音声を対象にした二波形分離問題の解法

6.2.1 はじめに

ここでは、本章で提案した音の分離抽出における聴覚の計算の方略を、実音声（母音）を対象にした二波形分離問題に展開し、計算の方略の正当性を示す。これは、二波形分離モデルが音声認識のフロントエンドとして応用できることも示す。そこで、本節では実音声と雑音の混ざった二波形分離問題に的を絞り、本モデルの分離精度を評価する。特に、雑音下での単母音・連続母音の分離抽出の精度を制約条件のいくつかを省略した場合について評価することで、本モデルの有効性を示す。また、本モデルが二重母音中から目的の母音を分離抽出できることも示す。

ここで、分離抽出の対象音は有声音に限定しているため、本節で取り扱う二波形分離問題には、第３章で提案した二波形分離問題の解法をそのまま適用できる。つまり、表5.1の制約条件をすべて利用すればよいということである。

6.2.2 二波形分離モデルの性能評価実験

本モデルが、雑音下における実音声の分離抽出において、どの程度正確に混合信号 $f(t)$ から望みの音声 $f_1(t)$ を分離抽出できるかを評価するために、次の三種類の評価実験を行う。

1. 雑音下の単母音の分離抽出
2. 雑音下の連続母音の分離抽出
3. 二重母音の中からの単母音の分離抽出

特に、(1) と (2) では基本周波数の時間的変動や、調音結合の有無にかかわらず本モデルが正確に目的音を分離抽出できることを、(3) では複数音声が存在する場合でも本モデルが正確に目的音を分離抽出できることを示すことが狙いである。また、評価に利用する実験データとして、ATR 音声データベースデータセット [Takeda *et al.*, 1988] にある男性２名

(mau, mht) と女性２名 (fkn, fsu) の単母音 (/a/, /i/, /u/, /e/, /o/) と連続母音 (/aoi/) を利用する。また、雑音については、帯域幅 6 kHz で帯域制限されたランダム雑音とピンク雑音を利用する。

次に、モデルの分離精度の評価尺度について説明する。音声認識のフロントエンドとして ASA のアプローチを取った音源分離モデルの研究 [Okuno *et al.*, 1997 ; 柏野ら, 1996a] では、モデルの分離精度の評価として認識率のみを利用している。しかし、純粋に ASA に基づくモデルの性能を評価するのであれば、「雑音をどの程度分離 (除去) できるのか」と「その効果により認識率がどの程度向上するのか」を議論する必要があると思われる。そこで本論文では、認識率の議論は一切せずに、モデルの分離精度を評価尺度として、$f_1(t)$ を信号、$f_1(t)$ と $\hat{f}_1(t)$ の差を雑音とみなした時間領域における SNR (式 (4.2)) を利用する。この評価尺度を用いることで、二波形の瞬時振幅だけでなく瞬時位相も正確に分離でき、かつ正確に波形レベルに復元できることを示すことができる。

次に、本モデルで利用した制約条件の有効性を考察するために、第４章で利用した三つの条件：

Condition 1 Comb filter による調波成分抽出＋ Kalman filter で求めた $C_{k,0}(t)$ と $D_{k,0}(t)$ の利用

Condition 2 Comb filter による調波成分抽出

Condition 3 処理なし (分析合成系による全域通過)

の比較も行う。ここで、Condition 1 は、制約条件 2.2 のなめらかさを省略した場合、Condtion 2 は制約条件 2 の漸近的変化を省略した場合、Condition 3 は、すべての制約条件を省略したものである。

評価実験１：雑音下の単母音の分離抽出

評価実験１では、表 6.1の 1. に示す $f_1(t)$ と $f_2(t)$ の SNR を 5 dB から 25 dB まで 5 dB 刻みに変化させた、合計 200 個 (5 SNR × 4 話者× 5 母音× 2 雑音) の混合信号 $f(t)$ を利用する。

例えば、図 6.1 (a) に示すような $f_1(t)$ (話者 mau の母音/a/) に、SNR が 15 dB のピンク帯域雑音を付加したとき、図 6.1 (b) に示すような混合信号 $f(t)$ となる。本モデルは、図 6.1 (c) に示すように $f(t)$ から $f_1(t)$ の基本周波数を推定し、図 6.1 (d) に示すように、混合信号 $f(t)$ から $\hat{f}_1(t)$ を 25.7 dB の精度で分離抽出できる。

表 6.1: 実験データ.

Sim. No.	$f_1(t)$	$f_2(t)$
1	/a/, /i/, /u/, /e/, /o/ (mau, mht, fkn, fsu)	ピンク帯域雑音 or ランダム帯域雑音
2	/aoi/ (mau, mht, fkn, fsu)	ピンク帯域雑音 or ランダム帯域雑音
3	/a/, /i/, /u/, /e/, /o/ (mau or fkn)	/aoi/ (fsu or mht)

次に、本モデルと三つの条件（Condition 1、2、3）の比較を行ったところ、図 6.2の結果を得た。図 6.2 (a), (b) はそれぞれ、$f_2(t)$ がピンク帯域雑音とランダム帯域雑音のときの分離精度を示す。また、図中の棒グラフは分離精度の平均（話者と母音の数で平均をとったもの）を、縦棒は標準偏差を示す。この図から、本モデルを利用した場合の分離精度が他の三つの条件よりも良好であることがわかる。Condition 1 との比較では、なめらかさ（制約条件 2.2）の制約を利用したことによる分離精度の向上を確認できる。本モデルと Condition 1、および Condition 2 の比較では、同一周波数領域に二波形の成分が存在する際、位相情報を利用したことによる分離精度の向上を確認できる。本モデルと Condition 3 の比較では、本モデルの分離精度の向上（雑音除去能力）を求めることができる。この結果、$f(t)$ の SNR が 5 dB（最悪 SNR）のとき、$\hat{f}_1(t)$ の分離精度が、ピンク帯域雑音で 9.2 dB、ランダム帯域雑音で 4.3 dB、改善されたことがわかる。

評価実験２：雑音下の連続母音の分離抽出

評価実験２では、表 6.1の 2. に示す $f_1(t)$ と $f_2(t)$ の SNR を 5 dB から 25 dB まで 5 dB 刻みに変化させた、合計 40 個（5 SNR × 4 話者× 1 連続母音× 2 雑音）の混合信号 $f(t)$ を利用する。

例えば、図 6.3 (a) に示すような $f_1(t)$（話者 mau の母音/aoi/）に、SNR が 15 dB のピンク帯域雑音を付加したとき、図 6.3 (b) に示すような混合信号 $f(t)$ となる。本モデルは、図 6.3 (c) に示すように $f(t)$ から $f_1(t)$ の基本周波数を推定し、図 6.3 (d) に示すように、混合信号 $f(t)$ から$\hat{f}_1(t)$ を 17.2 dB の精度で分離抽出できる。

次に、本モデルと三つの条件（Condition 1、2、3）の比較を行ったところ、図 6.4の結果を得た。図中の平均と標準偏差は図 6.2と同じ方法で計算したものである。この図から、

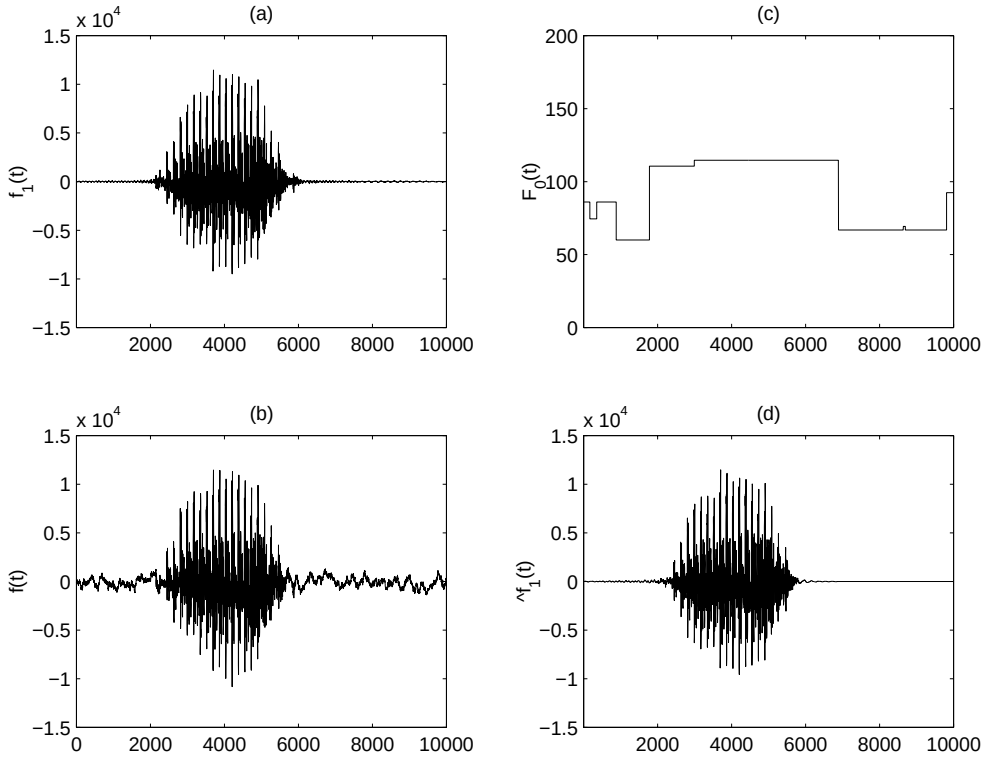

図 6.1: 評価実験１の分離例：(a) 原信号 $f_1(t)$（mau /a/）、(b) 混合信号 $f(t)$、(c) 基本周波数 $F_0(t)$、(d) 分離抽出された信号$\hat{f}_1(t)$

本モデルを利用した場合の分離精度が他の三つの条件よりも良好であることがわかる。この結果、$f(t)$ の SNR が 5 dB（最悪 SNR）のとき、$\hat{f}_1(t)$ の分離精度が、ピンク帯域雑音で 7.3 dB、ランダム帯域雑音で 5.7 dB、改善されたことがわかる。

評価実験３：同時音声からの単母音の分離抽出

評価実験３では、表 6.1の 3. に示す $f_1(t)$ と $f_2(t)$ の SNR を 5 dB から 25 dB まで 5 dB 刻みに変化させた、合計 50 個（5 SNR × 2 話者× 5 母音× 1 妨害音声）の混合信号 $f(t)$ を利用する。

例えば、図 6.5 (a) に示すような $f_1(t)$（話者 mau の母音/a/）に、SNR が 5 dB の $f_2(t)$（話者 fsu の連続母音/aoi/）を付加したとき、図 6.5 (b) に示すような混合信号 $f(t)$ となる。本モデルは、図 6.5 (c) に示すように $f(t)$ から $f_1(t)$ の基本周波数を推定し、図 6.5 (d) に示すように、混合信号 $f(t)$ から$\hat{f}_1(t)$ を 10.2 dB の精度で分離抽出できる。

次に、本モデルと三つの条件（Condition 1、2、3）の比較を行ったところ、図 6.6の結果を得た。図中の平均と標準偏差は図 6.2と同じ方法で計算したものである。この図から、

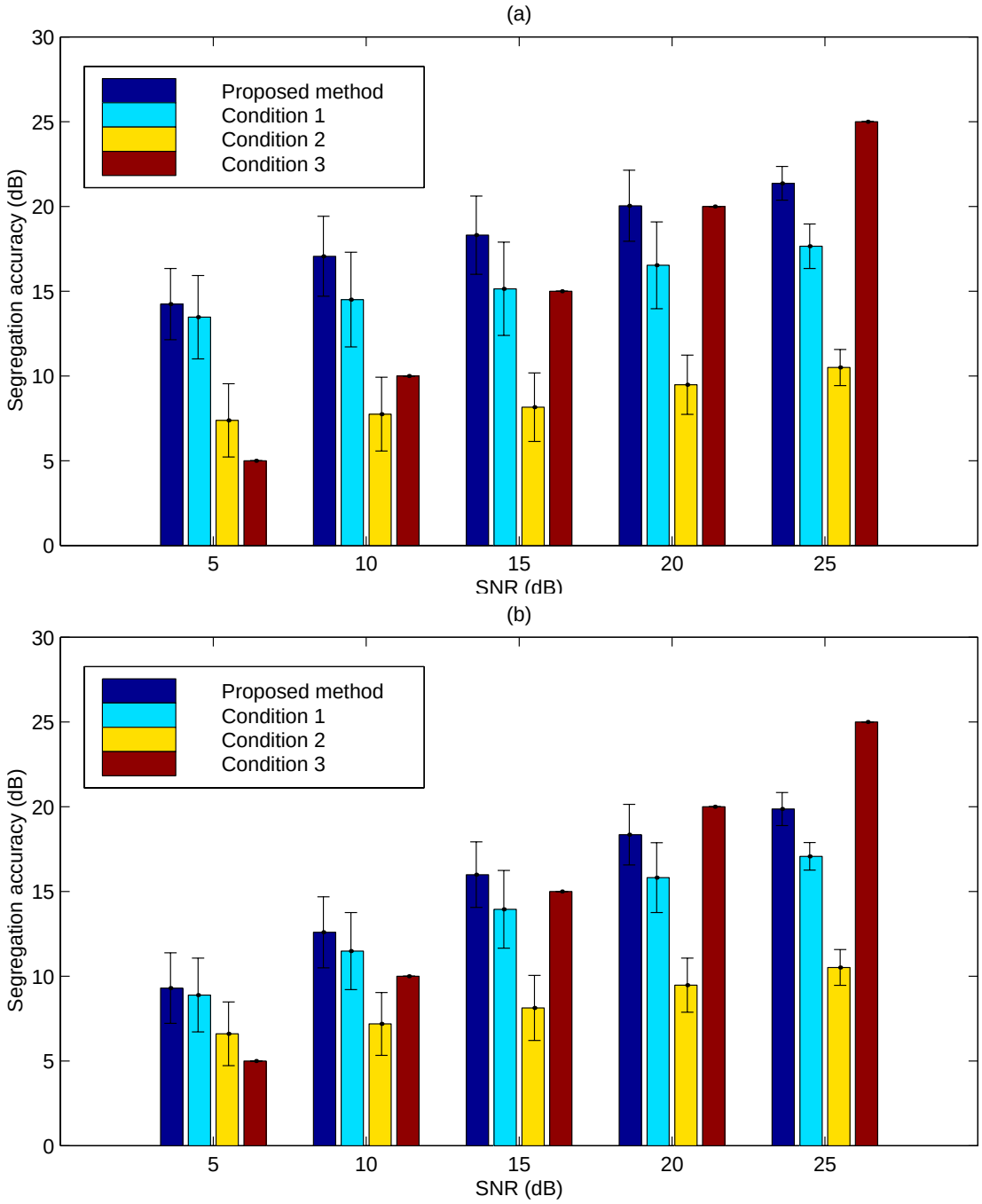

図 6.2: 評価実験 1 の分離精度の比較：(a) ピンク帯域雑音の場合, (b) ランダム帯域雑音の場合

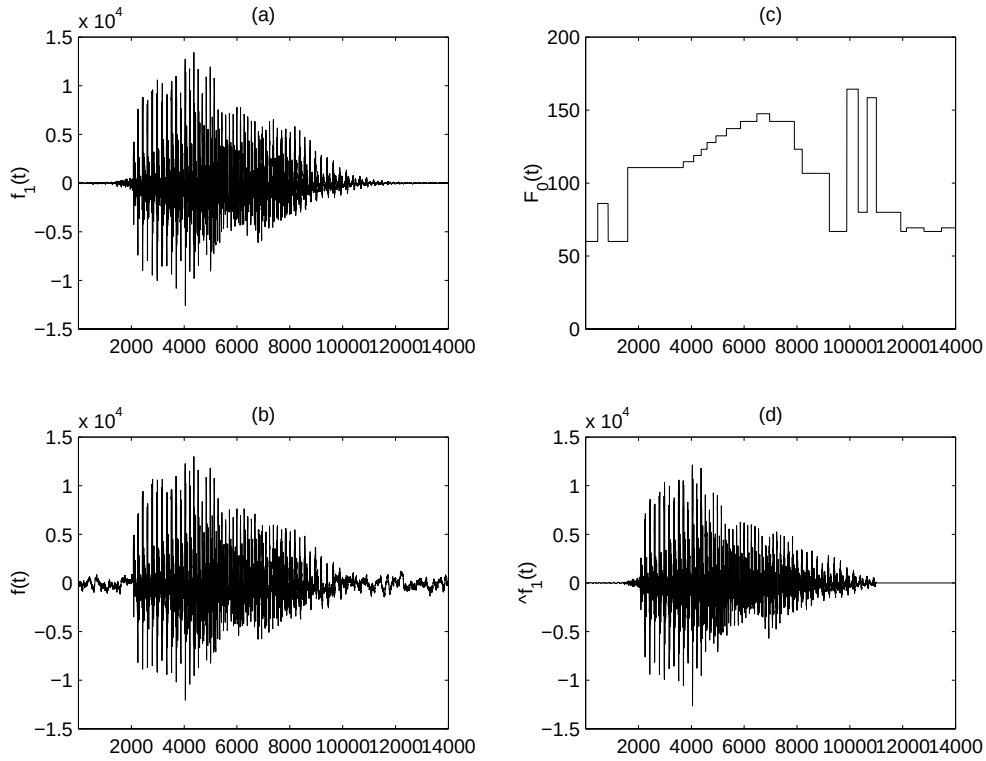

図 6.3: 評価実験２の分離例：(a) 原信号 $f_1(t)$（mau /aoi/）、(b) 混合信号 $f(t)$、(c) 基本周波数 $F_0(t)$、(d) 分離抽出された信号$\hat{f}_1(t)$

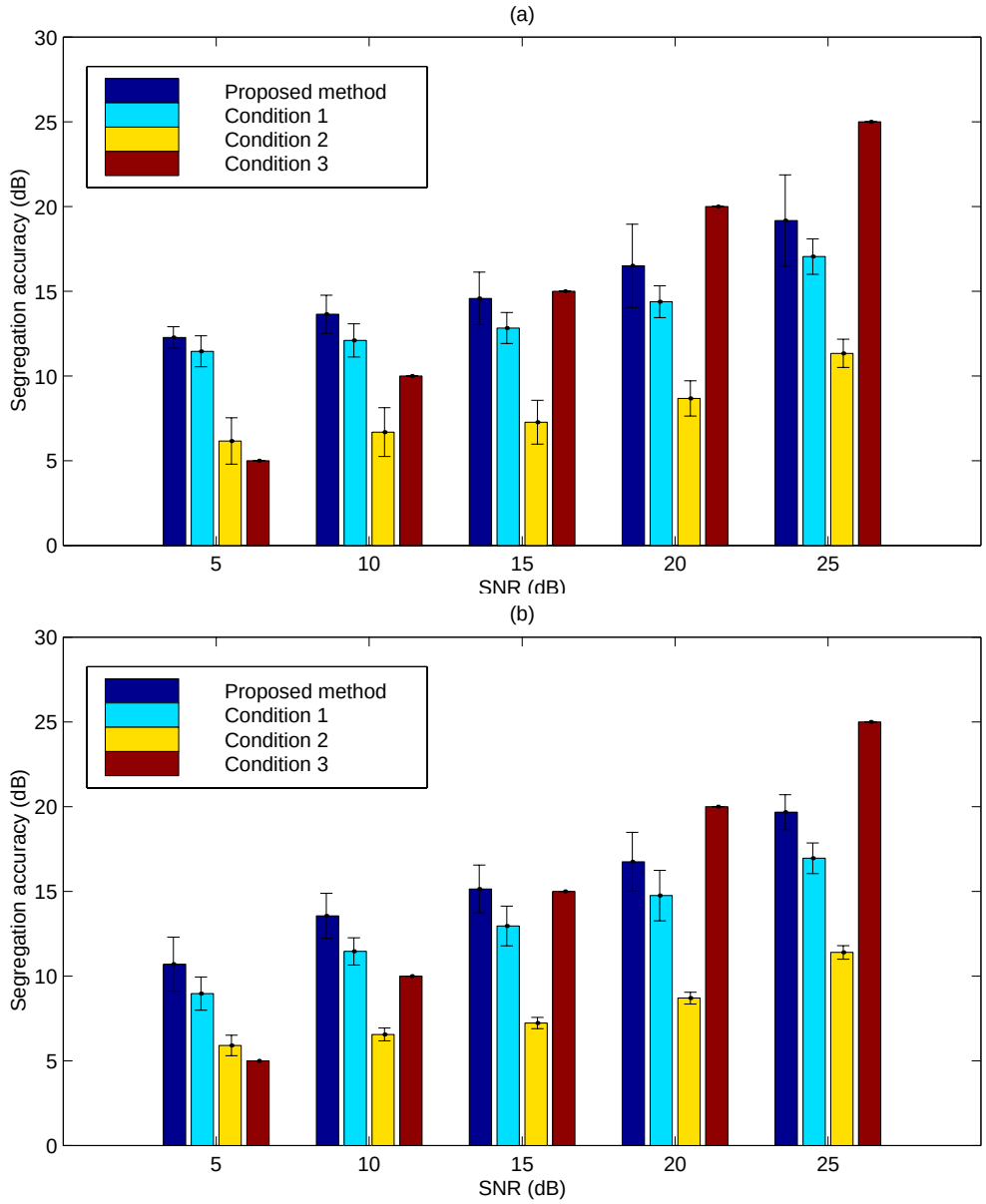

図 6.4: 評価実験 2 の分離精度の比較：(a) ピンク帯域雑音の場合、(b) ランダム帯域雑音の場合

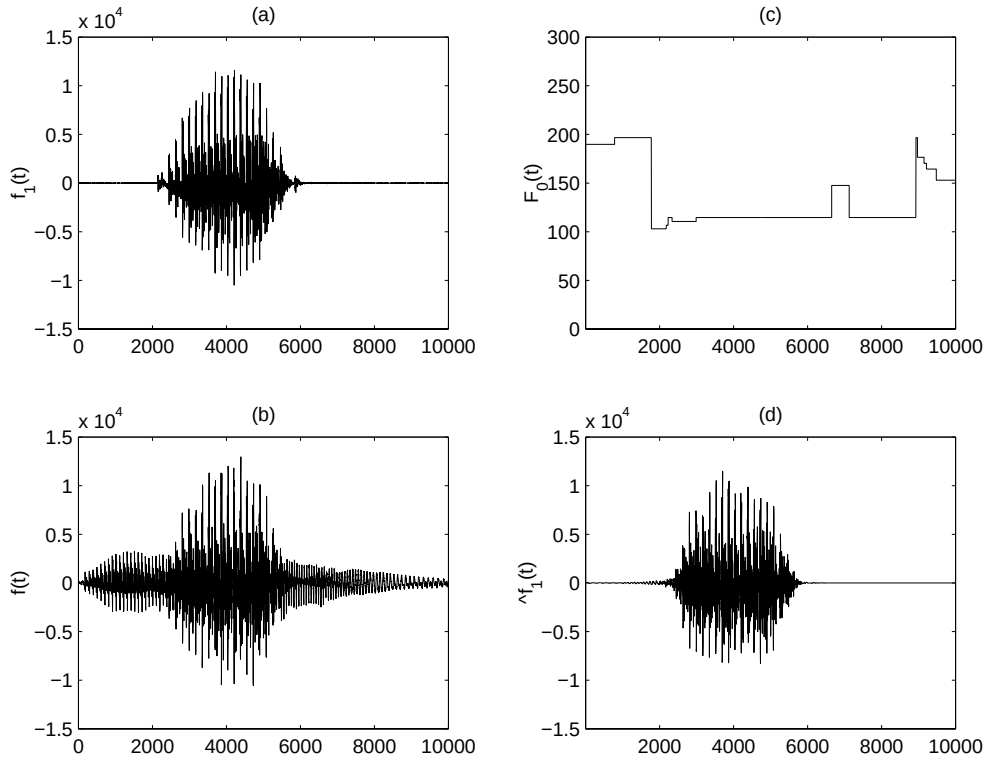

図 6.5: 評価実験３の分離例：(a) 原信号 $f_1(t)$（mau /a/）、(b) 混合信号 $f(t)$（mau /a/ + fsu /aoi/）、(c) 基本周波数 $F_0(t)$、(d) 分離抽出された信号$\hat{f}_1(t)$

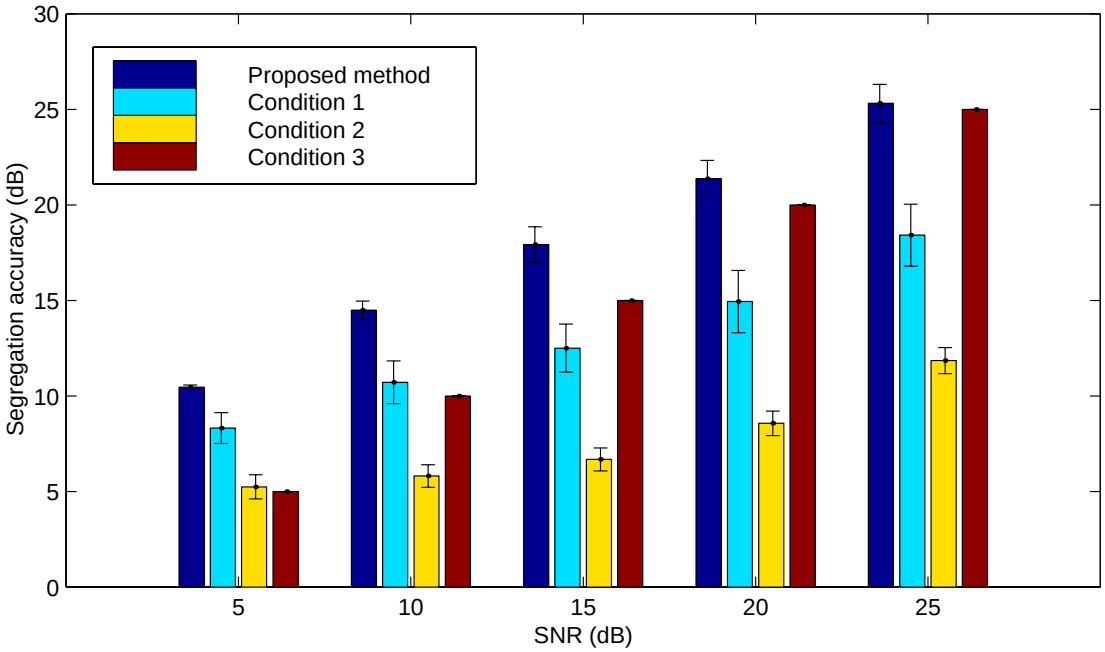

図 6.6: 評価実験３の分離精度の結果

本モデルを利用した場合の分離精度が他の三つの条件よりも良好であることがわかる。この結果、$f(t)$ の SNR が 5 dB（最悪 SNR）のとき、$\hat{f}_1(t)$ の分離精度が 5.2 dB 改善されたことがわかる。

6.2.3 考察

図 6.2, 図 6.4, 図 6.6の結果から、本モデルの有効性が示された。すなわち、実音声を対象にした二波形分離問題においても、本論文で提案した計算の方略（制約条件すべて）の正当性が示された。但し、SNR が最良時（25 dB 以上）のとき、混合信号を原信号と見なした場合（Condition 3 に対応）の分離精度と大差ないか、もしくは若干低下する結果となった。これは、二波形分離モデルにおける瞬時振幅、瞬時位相、基本周波数の時間変動を R次の区分多項式で近似する際、計算量の削減のために $R = 1$ としたことに起因するものであり、実装上で生じた問題である。そのため、これは、多項式近似の次数を高くすることで改善される。

次に、評価実験１と２の結果から、$f_2(t)$ の種類に依存して本モデルの分離精度に差が生じていることがわかる。これは、本モデルで採用した定 Q gammatone filterbank の構成に起因する。各分析フィルタ形状は一定の Q をもつため、ピンク帯域雑音のフィルタ通過成

分のパワーはおおよそ均一に分散し、ランダム帯域雑音のフィルタ通過成分のパワーは高域側に集中する。一方、調波成分は低域側では比較的安定して調波関係を満たすが、高域側では調波関係を正確に満たさない可能性がある。この相乗効果により、高域において誤った調波成分に付加された、帯域雑音成分の未抽出成分の影響が分離精度の低下を招いているものとも考えられる。

最後に、評価実験３において、本モデルが二重母音の分離抽出問題にも適用可能であることがわかる。また、本モデルと Condition 1、および Condition 2 を比較すると、同一周波数領域に二波形の成分が存在する際、位相情報を利用したことによる分離精度の向上が確認される。

以上の考察から、本節における二波形分離問題の検討は、本論文で提案された計算の方略により導出された解法の正しさを実証しただけでなく、二波形分離モデルが雑音にロバストな音声認識のフロントエンドとしての適用にも期待できる結果を示した。

6.3 共変調マスキング解除を想定した二波形分離問題の解法

6.3.1 はじめに

聴覚系の周波数選択性の研究において、マスキングの現象を説明するモデルとして、マスキングのパワースペクトルモデル [Patterson and Moore, 1986] が広く受け入れられている。このモデルでは、聴取者が、背景雑音中で特定の中心周波数をもつ正弦波信号を検知しようとするとき、信号周波数付近で中心周波数をもち、信号対雑音比の最も高くなる単一の聴覚フィルタの出力を利用するものと仮定している。また、刺激は長時間パワースペクトルとして表現されており、信号のマスキングしきい値は、聴覚フィルタを通過する雑音の量によって決定されるものと仮定している。この一連の仮定により、パワースペクトルモデルは同時マスキングといった多くの現象をよく説明できるが、成分音間の相対位相やマスカーの短時間変動を無視しているため、説明できないマスキング現象もいくつか存在した。

Hall（1984）らは聴覚フィルタ間の比較によって、振幅包絡が変動する雑音にマスクされた正弦波信号の検出が容易になるという可能性を示した [Hall and Fernandes, 1984]。このような検知能力の向上が生じるための決定的な条件は、異なる周波数帯域間で振幅包絡の変動が一致しているか、あるいは相関があるということであった。Hall らはこの異なる周波数帯域間の振幅包絡の一致を “共変調” と呼び、これによる検知能力の向上、つまりマスキングの解除を共変調マスキング解除（Co-modulation Masking Release: CMR）と

呼んだ。この現象については、多くの聴覚心理実験 [Hall and Grose, 1988 ; Moore, 1997 ; Moore, 1992 ; Willen,*et al.*, 1992] が行われており、同様の結果が得られている。

一方、共変調マスキング解除については、もう一つの見解がある。これは、くぼみ聴取モデル（dip-listening model）と呼ばれ、マスカーの振幅包絡の変動が極小になるときと純音の包絡が最大になるところが重なったときに、純音を検知できるというものである [Buus, 1985]。この見解についてもいくつかの追試実験が行われているが、どちらの見解が正しいのかはまだ明らかになっていない。最近の聴覚心理実験の報告を見る限りは、Hall らの見解を支持している研究者が多いようである。本研究では、前者の知見を支持する。

以上、二つの見解を踏まえると、共変調マスキング解除がどちらの手がかりを利用して行われているのか分からないが、単一の手がかりやメカニズムに基づいているのではないということは確かである。しかし、共変調マスキング解除が起こるための条件が上記に示したように知られているにもかかわらず、これらの条件を利用した計算モデルはほとんど報告されていない。

そこで、本節では、前章で提案した音の分離抽出における聴覚の計算の方略を共変調マスキング解除を想定した二波形分離問題に展開し、計算モデルの実装を試みる。CMR の分離問題において、分離抽出の対象となる音は純音であり、もう一方の音はマスカーとなる。従って、第２章の考察から、共変調マスキング解除を想定する二波形分離問題では四つの発見的規則のうち効果を発するのは発見的規則 (ii) と (iv) である。また、分離抽出の対象音が純音の場合、第４章の十分性の検証から、表 5.1の制約条件の実装は図 4.2に示すパラメータ決定方法で実現できる。

本節では、以上の設定を踏まえ、同時マスキング現象の説明に利用されてきたマスキングのパワースペクトルモデルと、第３章で提案した二波形分離モデルの二つのモデルを利用し、これに二つのモデルの結果を選択する処理を付加することで、CMR の計算モデルを提案する。

6.3.2 CMR の計算モデル

CMR の計算モデルを図 6.7に示す。本モデルは、二つのモデル（A, B）と選択処理で構成される。また、$f_1(t)$ を純音（正弦波信号）、$f_2(t)$ をその純音の周波数を中心周波数とする二種類のマスカー (ランダム帯域雑音と振幅変調されたランダム帯域雑音) とし、$f_2(t)$ が存在している状態で $f_1(t)$ が加算される状況（同時マスキング）を想定している。本モデルはこの混合信号 $f(t)$ だけを受音でき、二つのモデル（A, B）を用いて、純音 $f_1(t)$ を分離抽出する。モデル A は、二波形分離モデルであり、モデル B はマスキングのパワースペクトル

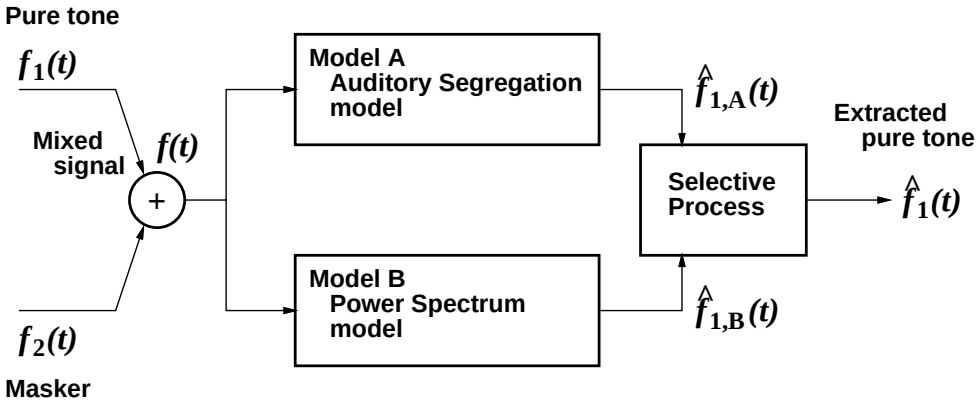

図 6.7: CMR の計算モデル

モデル [Patterson and Moore, 1986] である。CMR の計算モデルでは、これら二つのモデルが並行に動作し、それぞれ、マスクされた信号から純音を分離抽出する。ここで、モデル A によって分離抽出された純音を$\hat{f}_{1,\mathrm{A}}(t)$、モデル B によって分離抽出された純音を$\hat{f}_{1,\mathrm{B}}(t)$とする。最後に、選択処理部は、二つのモデルにより分離抽出された$\hat{f}_{1,\mathrm{A}}(t)$ と$\hat{f}_{1,\mathrm{B}}(t)$ のうちマスキングしきい値の低い方を選択し、これを計算モデルで分離抽出された純音$\hat{f}_1(t)$ とする。この計算モデルの根本的な考えは、Hall らの実験結果 [Hall and Fernandes, 1984] において、マスカー帯域幅が単一の聴覚フィルタの帯域幅（1 ERB）を越えるか越えないかでマスキングの傾向が異なっていたことに由来する。言い換えると、Hall らの実験結果において、マスカー帯域幅が 1 ERB を越えないときに、単一の聴覚フィルタの出力を手がかり（モデル B）として説明でき、マスカー帯域幅が 1 ERB を越えるときに、聴覚フィルタ群の出力を手がかり（モデル A）として説明できるというものである。

6.3.3 モデル A: 二波形分離モデル

モデル A で採用する二波形分離モデルは、図 3.2で提案したものである。ここで利用する分析フィルタ群は、$f_0 = 1$ kHz、$K = 128$、$b_f = 1.019$、解析可能な帯域幅を 10 Hz〜10 kHz で設計された聴覚フィルタ群であり、詳細については第３章を参照されたい。また、二波形分離アルゴリズムは、図 4.2に示すアルゴリズムを利用する。

6.3.4 モデル B: パワースペクトルモデル

Patterson と Moore によって提案されたマスキングの “パワースペクトルモデル” [Patterson and Moore, 1986] では、聴取者が背景雑音から正弦波信号を検知するときに、信号

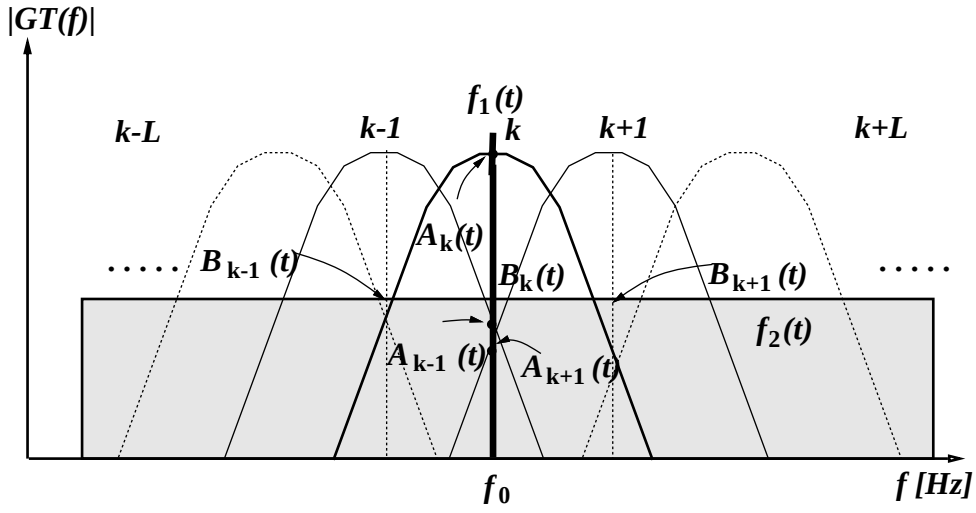

図 6.8: 隣接する聴覚フィルタ出力における純音 $f_1(t)$ とマスカー $f_2(t)$ の特性

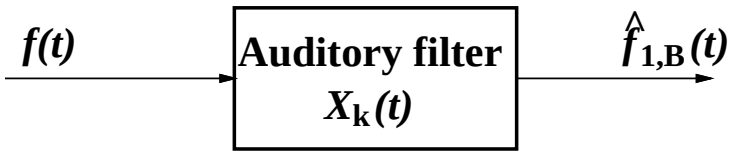

図 6.9: モデル B: パワースペクトルモデル

周波数近傍の中心周波数をもち、信号対雑音比が最も高い単一の聴覚フィルタの出力を利用するものと仮定した。従って、雑音の成分のうち、このフィルタを通過する成分だけが信号のマスキングに影響を及ぼすものと考えることができる。また、信号のしきい値は聴覚フィルタを通過する雑音の量によって決定される。

本論文では、このパワースペクトルモデルを図 6.9に示すモデルで構成する。ここで、$X_k(t)$ はモデル A で構成された聴覚フィルタ群のうちの一つの聴覚フィルタの出力である。また、聴覚フィルタは、中心周波数が 1 kHz、帯域幅が 1 ERB の gammatone フィルタで構成される。このモデルでは、混合信号 $f(t)$ に対し、単一聴覚フィルタ（帯域幅が 1 ERB の帯域通過フィルタ）を通過した出力を、分離抽出された純音 $\hat{f}_{1,\mathrm{B}}(t)$ とする。

ここで、純音の分離抽出に関するモデル A とモデル B の決定的な違いは、図 6.8の単一フィルタ出力（例えば、中心周波数が f_0のとき）において、その出力をそのまま利用して純音を分離抽出するか、あるいは隣接するフィルタ出力も利用して純音を分離抽出するかということである。

6.3.5 シミュレーション

共変調マスキング解除 (CMR)

Hall らは、共変調マスキング解除の実験の一つで、1 kHz、400 msec の正弦波信号のしきい値をスペクトルレベルを一定に保った雑音マスカーの帯域幅の関数として測定した [Hall and Fernandes, 1984 ; Moore, 1997]。また、マスカーの中心周波数は 1 kHz であり、次のような二種類のマスカーが用いられた。

- ランダム帯域雑音：振幅は不規則にかつ異なる周波数領域において独立に変動する。
- 振幅変調されたランダム帯域雑音：ランダム帯域雑音であるが、ランダム帯域雑音の振幅を不規則なゆっくりとした速度で変調（50 Hz の低域通過フィルタリング）したものである。振幅変動は異なる周波数領域において等しい。

この二種類のマスカーを用いて正弦波信号の検知能力を測定したところ、図6.10に示す結果が得られた。図中の点 R はランダム帯域雑音の場合の信号のしきい値を表し、点 M は振幅変調された雑音の場合の信号のしきい値を表している。この結果、帯域雑音の帯域幅が、この中心周波数での聴覚フィルタの帯域幅（約 130 Hz）を越えない場合、いずれの帯域雑音についてもマスキング量が増加している。一方、この帯域幅を越える場合、ランダム帯域雑音ではマスキング量が変動しないのに対し、振幅変調されたランダム帯域雑音の場合、マスカー帯域幅の増加に従って、マスキング量が減少している。この結果から、Hall らは、異なる聴覚フィルタ間の比較によって、聴取者は信号検出能力を高めることができることを示し、この現象を共変調マスキング解除（CMR: Co-modulation Masking Release）と呼んだ。この実験では、共変調マスキング解除量は最大約 10 dB であった。

本論文では、Hall らの実験と等価な条件を考慮し、本モデルが CMR の特性を模擬することを検証するため、次のような計算機シミュレーションを行う。

モデル A のシミュレーション

実験データは、Hall らの実験と等価な条件を考慮するため、サンプリング周波数 20 kHz、周波数を 1 kHz、呈示時間を 400 msec、振幅を一定とした正弦波信号 $f_1(t)$ と $f_1(t)$ の周波数を中心周波数とした二種類のマスカー $f_2(t)$(ランダム帯域雑音と振幅変調されたランダム帯域雑音) を用意した。ここで、$f_{21}(t)$ はランダム帯域雑音であり、ある乱数の種を設定することで生成される白色雑音を基に、これを帯域制限することで得られる。また、$f_{22}(t)$ は振幅変調されたランダム帯域雑音であり、変調速度が 50 Hz、変調度が 100 %で $f_{21}(t)$ を振幅変調（50 Hz の低域通過フィルタリング）したランダム帯域雑音である。このとき、

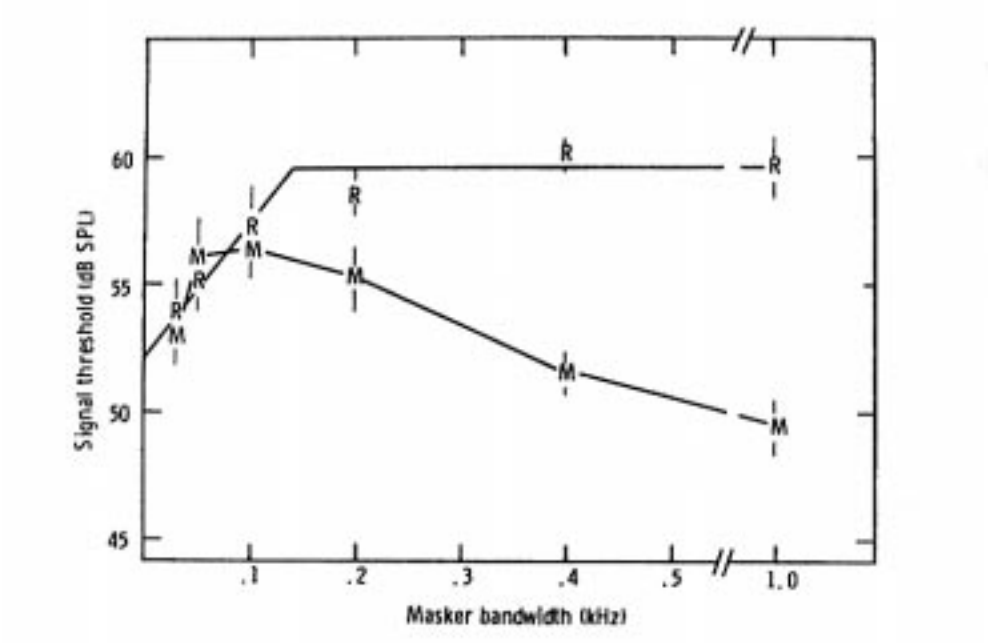

図 6.10: CMR の実験結果（Hall et al., 1984）

$f_2(t)$ のパワーが$\sqrt{f_{21}(t)^2/f_{22}(t)^2} = 1$ となるように調整され、$f_1(t)$ と $f_2(t)$ の SNR (signal to noise ratio) は−6.61 dB であった。これらの実験データを図 6.11（左側）に示す。ここで、各混合信号は $f_R(t) = f_1(t) + f_{21}(t)$, $f_M(t) = f_1(t) + f_{22}(t)$ であり、それぞれ Hall らの実験データで用いられた点 R、点 M の刺激に対応する。刺激は、開始時刻を変化させた純音 $f_1(t)$ を 10 個、乱数の種（5 種類）を変化させて作成した二種類のマスカーをそれぞれ 5 個とし、合計 50 個の混合信号を用意した。このときの混合信号の一例を図 6.11（右側）に示す。ここで、いずれの混合信号においても、純音 $f_1(t)$ は視覚的にマスカーに埋もれていたが、聴覚的には $f_M(t)$ で純音を容易に検知でき、$f_R(t)$ で純音を検知することが困難であった。

次に、Hall らの実験と等価なシミュレーション条件を考える際、CMR で利用する手がかりの幅を制御するために聴覚フィルタ間の帯域幅を知る必要があるが、この実験において人間がどの程度の幅の聴覚フィルタ間の手がかりを利用したか分からない。そのため、本研究では、CMR を起こすために与えた手がかりの帯域幅（マスカー帯域幅）と手がかりを扱える帯域幅（聴覚フィルタ間の帯域幅）を等価と考える。従って、ここでは、Hall らによるマスカー帯域幅の関数としてマスキングしきい値を測定した方法を、マスカー帯域幅をあらかじめ広めに固定（1 kHz）しておき、隣接する聴覚フィルタの参照数 L（利用する聴覚フィルタ間の全帯域幅に対応）の関数として、しきい値を測定することと見なす。

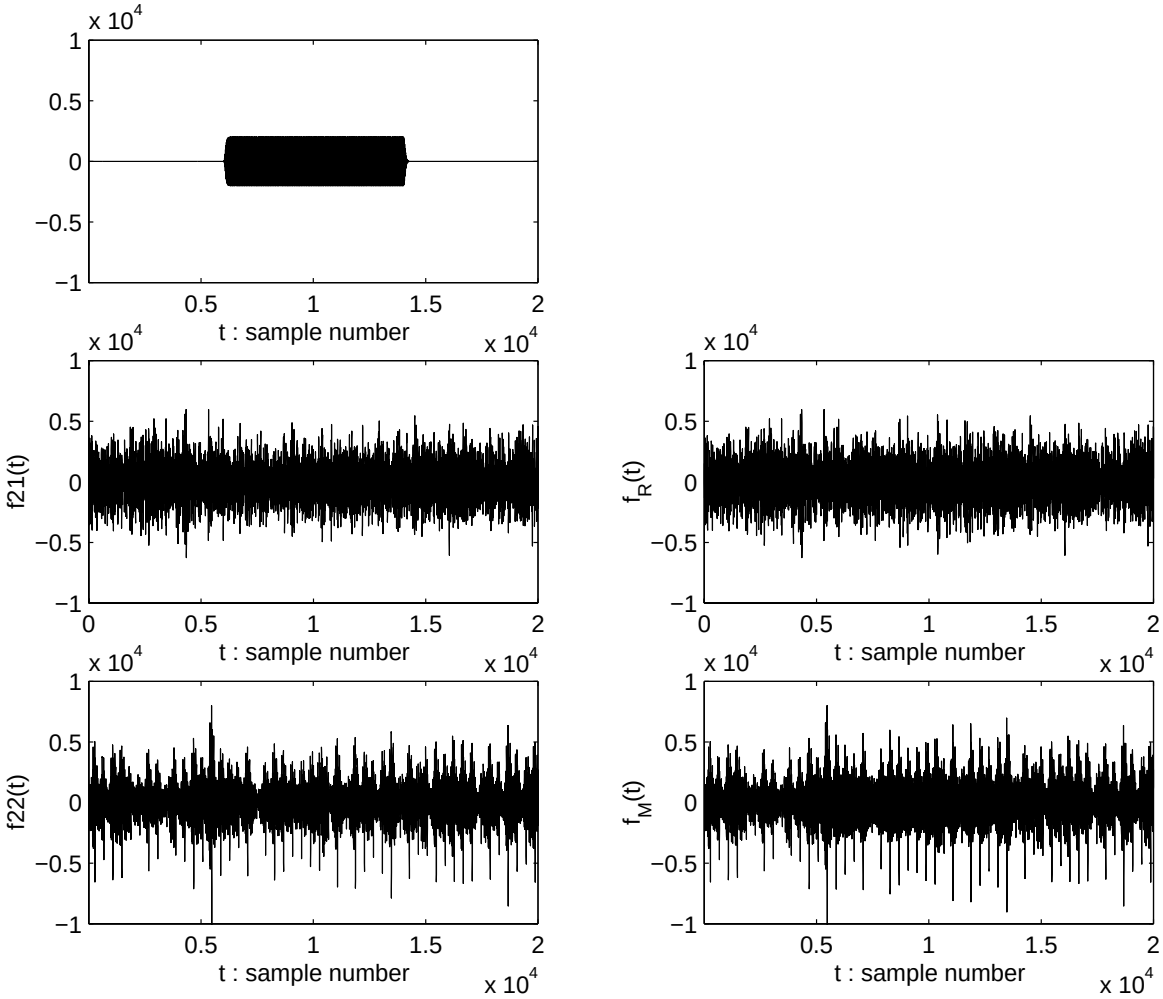

図 6.11: 刺激：（左上）原信号：純音 $f_1(t)$, （左中）ランダム帯域雑音 $f_{21}(t)$, （左下）振幅変調されたランダム帯域雑音 $f_{22}(t)$, （右上）混合信号 $f_R(t)$, （右下）混合信号 $f_M(t)$

また、しきい値は分離抽出された$\hat{f}_{1,A}(t)$ の SN 比（分離精度）と見なし、マスキングからの解除量をちょうど SN 比の改善量に対応させる。このとき、入力位相 $\theta_{2k}(t)$ は、隣接する聴覚フィルタの参照数 L の関数として求められた$\hat{\hat{B}}_k(t)$ によって一意に決定される。ここで、参照数は $L = 1, 3, 5, 7, 9, 11$ とし、これに対応する帯域幅はそれぞれ 207, 352, 499, 648, 801, 958 Hz である。

シミュレーション条件に従い、各混合信号についてシミュレーションを行った。このときの結果を図 6.12に示す。この図の縦軸は分離抽出された純音の SN 比の向上量を下向きに表し、横軸は、隣接する聴覚フィルタの参照数 L に対応した帯域幅を表している。また、図中の実線と縦棒はそれぞれ 50 個の混合信号に対して分離抽出された正弦波信号$\hat{f}_{1,A}(t)$ の

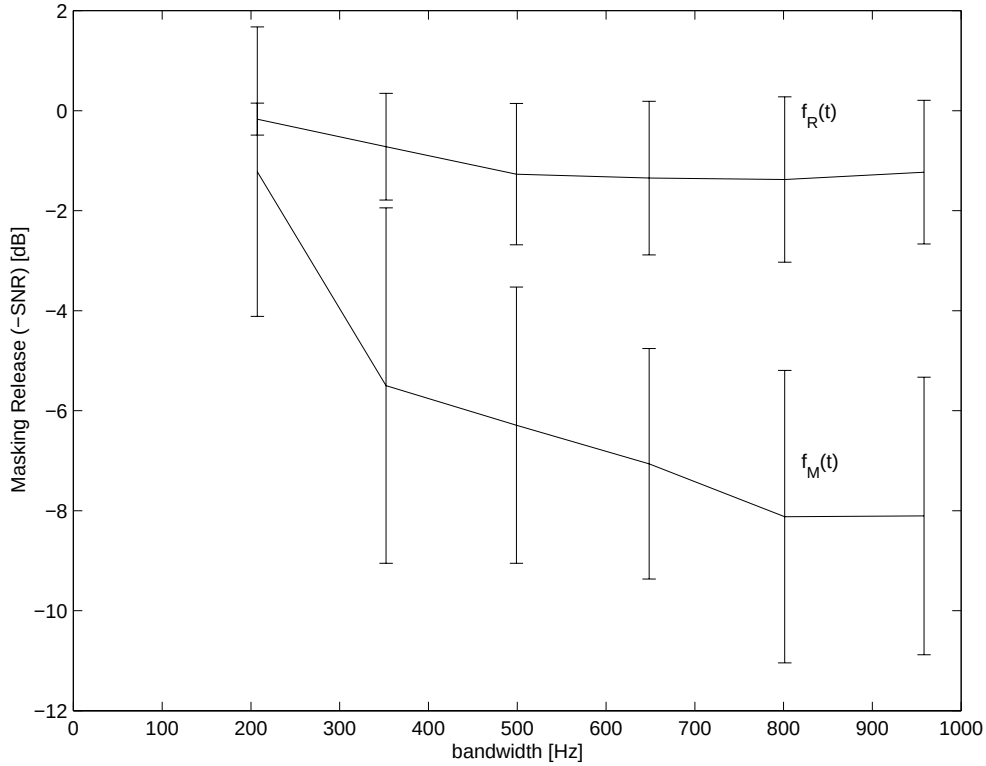

図 6.12: 隣接する聴覚フィルタの参照数 L に対応した帯域幅と $\hat{f}_{1,\mathrm{A}}(t)$ の分離精度（SNR）の関係

SN 比の平均値と標準偏差を表している。この結果、混合信号 $f_M(t)$ の場合、隣接する聴覚フィルタの参照数 L を増加させると、分離抽出された純音 $\hat{f}_{1,A}(t)$ の SN 比が向上する傾向が見られた。しかし、混合信号 $f_R(t)$ の場合、隣接する聴覚フィルタの参照数 L を増加させても、純音はほとんど抽出されず、SN 比はほとんど変わらなかった。従って、この結果は、マスカーの振幅成分が異なる周波数領域において同じ振幅変調パターンをもつとき、すなわち、マスカーの振幅包絡間の相関が高いとき、純音 $f_1(t)$ をより分離抽出しやすくなるという結果を示している。故に、この結果から、モデル A は、複数の聴覚フィルタ出力を利用し、マスカーの振幅包絡間の相関を手がかりにマスキング解除のメカニズムを模擬しているといえる。

モデル B のシミュレーション

実験データは、モデル A で利用したものと同様開始時刻を変化させた 10 個の純音 $f_1(t)$ を利用するが、二種類のマスカーについては、乱数の種（５種類）と帯域幅（９種類）を変化させて作成した 45 個とし、合計 450 個の混合信号を用意した。ここでマスカー帯域幅は、

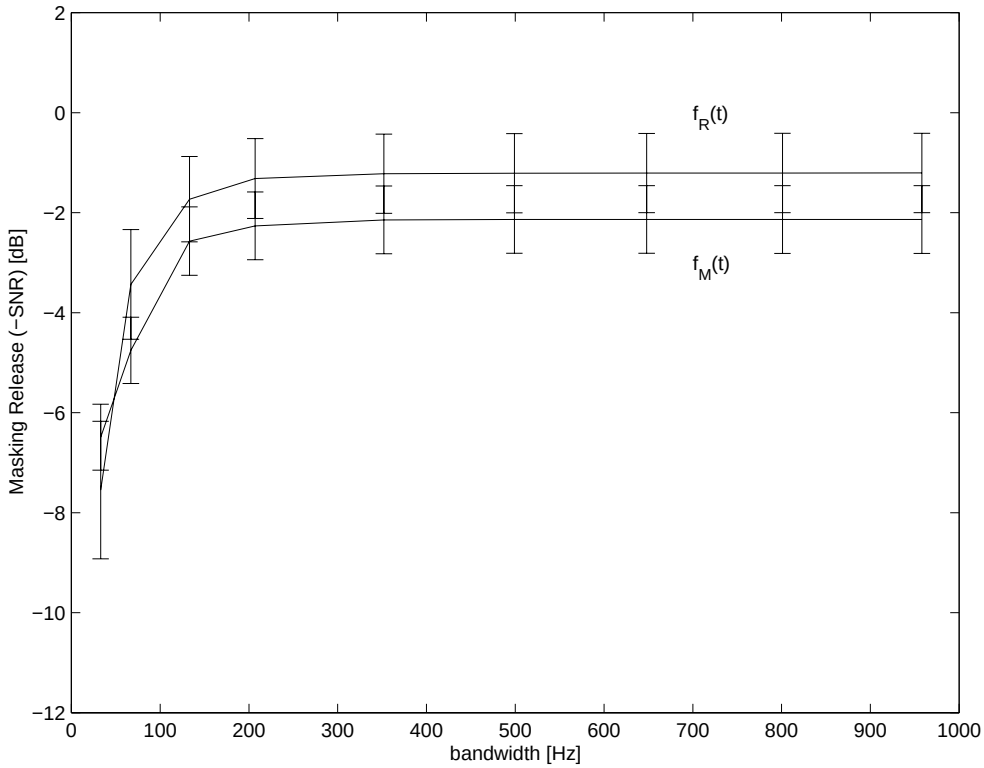

図 6.13: マスカー帯域幅と$\hat{f}_{1,\mathrm{B}}(t)$の分離精度（SNR）の関係

1/4 ERB, 1/2 ERB, 1 ERB に対応した 33, 67, 133 Hz の他、モデル A での $L = 1, 3, \cdots, 11$ に対応した 207, 352, 499, 648, 801, 958 Hz である。

モデル B では、Hall らの条件と同様、マスカー帯域幅の関数としてマスキングしきい値を測定する。また、モデル A の条件と同様に$\hat{f}_{1,B}(t)$ の SN 比をしきい値と見なす。

シミュレーション条件に従い、各刺激についてシミュレーションを行った。このときの結果を図 6.13に示す。この図の縦軸は分離抽出された純音の SN 比の向上量を下向きに表したものであり、横軸はマスカー帯域幅を表している。また、図中の実線と縦棒は、それぞれ SN 比の平均と標準偏差を表している。この結果、マスカーの種類に関係なく、マスカー帯域幅の増加とともにマスキングしきい値が変化していることがわかる。マスカー帯域幅が単一の聴覚フィルタの帯域幅に相当する 1 ERB を越えない場合、マスカー帯域幅の関数としてしきい値は増加しているが、マスカー帯域幅が 1 ERB を越える場合、しきい値はそれ以上増加せず一定になっている。

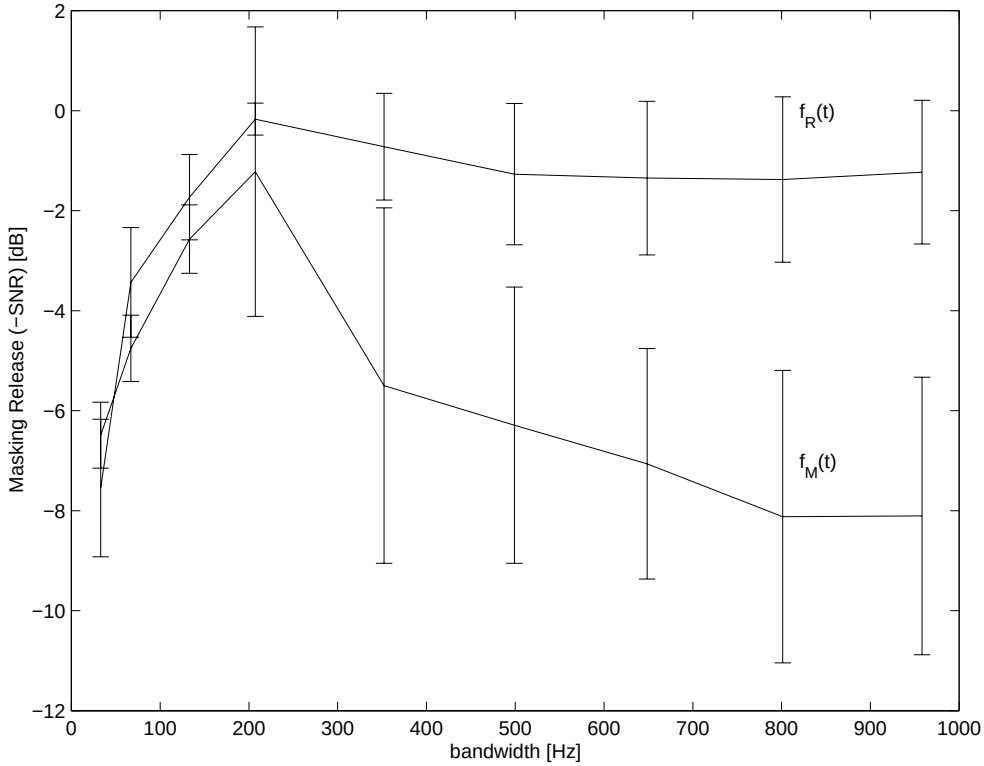

図 6.14: マスカー帯域幅と純音$\hat{f}_1(t)$の分離精度（SNR）の関係

6.3.6 CMR の計算モデルの特性

二つのモデルについて、シミュレーションを行った結果、モデル A ではマスカー帯域幅が 1 ERB を越えたとき、マスカーの振幅包絡間の変動の一致／不一致による共変調マスキング解除／マスキングの現象を模擬していることがわかる。また、モデル B ではマスカー帯域幅が 1 ERB を越えるまでマスキングしきい値が増加し、1 ERB を越えた後しきい値がそれ以上増加せず一定になるというマスキング現象を模擬していることがわかる。選択処理では、これら二つのモデルの処理結果のうち、マスキングしきい値の低いもの、言い換えると分離抽出された正弦波信号の SN 比の高いものを選択することで、図 6.12と図 6.13の結果から、図 6.14に示すような結果を得る。この特性は、図 6.10に示した Hall らの実験結果と類似した結果を示す。従って、本モデルは共変調マスキング解除の計算モデルと解釈できる。特に、共変調マスキング解除量は、Hall らの結果では最大約 10 dB であったのに対し、本モデルでは最大約 8 dB であった。

6.3.7 おわりに

本節では、本論文で提案した計算の方略を共変調マスキング解除を想定した二波形分離問題に展開することで、共変調マスキング解除の計算モデルを実現できた。このモデルは、二波形分離モデル（モデル A）とマスキングのパワースペクトルモデル（モデル B）の二つのモデルと、この二つのモデルの結果を選択する処理で構成された。マスカーから純音を検出するメカニズムについて、モデル A では、複数の聴覚フィルタの出力を手がかりにしているのに対し、モデル B では、単一の聴覚フィルタからの出力を手がかりにしている。Hall らによる共変調マスキング解除の実験を想定したシミュレーションを二つのモデルについてそれぞれ行った。モデル A では、マスカーの種類によってマスキングしきい値に変動があった。これは、ランダム帯域雑音の場合、マスカー帯域幅の増加に関係なくしきい値は変動しなかったものの、AM ランダム帯域雑音の場合、マスカー帯域幅の増加とともにマスキング解除が起こるという結果が得られた。モデル B ではマスカーの種類に関係なく、マスカー帯域幅の増加とともにマスキングしきい値が増加した。このしきい値は、マスカー帯域幅が 1 ERB を越えるまで増加したが、1 ERB を越えてからはそれ以上増加せず一定であった。この結果に対し、選択処理は二つのモデルの結果から分離抽出した純音のマスキングしきい値の低いものを選択することで、Hall らが示した CMR の結果と同様の傾向を示す特性が得られた。このとき共変調マスキング解除量は最大約 8 dB であった。

以上の結果から、本モデルが CMR の計算モデルと解釈できた。また、CMR の手がかりとして、規則 (iv) が有効であることも確認できた。

6.4 むすび

本章では、音の分離抽出における聴覚の計算の方略を (1) 実音声（母音）を対象とした二波形分離問題、(2) 共変調マスキング解除を想定した二波形分離問題に展開することで、本計算の方略がこれらの問題の解法を導出できることを実証した。(1) の結果から、音の分離抽出における聴覚の計算理論が音声認識のフロントエンドとして応用できることを示した。特に、雑音下での単母音・連続母音の分離抽出の精度を制約条件のいくつかを省略した場合について評価することで、本モデルの有効性を示すことができた。また、本モデルが二重母音中から目的の母音を分離抽出できることもできた。(2) の結果からは、本計算理論を展開することで、これまでに計算モデルが提案されていなかった CMR の計算モデルを実現できることを示した。また、Bregman によって提唱された発見的規則 (iv) が、数理工学的にも共変調マスキング解除で利用する手がかりとして有効であることが示された。

以上、二つの貢献できる側面を示したが、本論文で提案した音の分離抽出に関する聴覚の計算理論は、これ以外にも、カクテルパーティー効果のモデル化の手がかりや他のマスキング現象などのモデル化に向けても期待できるものと思われる。

第 7 章

結論

7.1 本論文で明かにされたことの要約

本論文では、音の分離抽出における聴覚の計算理論の構築を試みるために、

- 混合された信号から目的の原信号を求める信号分離問題を一意に解くためには音や環境に対する制約条件が必要である。
- 信号分離問題を聴覚の情景解析問題としてとらえ直せば、信号分離問題を一意に解くために必要な制約条件として聴覚が情景解析問題で利用している心理学的な制約条件を利用できる。

という考え方を採用した。そして、計算論的神経科学のアプローチで利用された方法論と同様、アルゴリズムの研究から計算理論の研究へと発展させることで、音の分離抽出における聴覚の計算理論の構築を試みた。従って、本研究では、聴覚系でどのような制約条件を設けることで目的の処理（不良設定問題を一意に解くこと）が可能なのかを、アルゴリズムより一段上のレベルで検討しなければならなかった。また、この制約条件を利用したアルゴリズムが計算理論から導かれたものになるためには、制約条件の必要十分性を示さなければならなかった。本研究では、心理学的・生理学的に意味のある制約条件を用いたアルゴリズムであり、かつ上記のようにアルゴリズムの研究から計算理論の研究へと発展する道筋が明確にあるアルゴリズムを「計算の方略」と呼んだ。ここで、計算の方略は、Marr の示した計算理論でいう「何を計算しているのか」を説明できるものであるが、その必要十分性が導かれることにより、「何故それをするのか？(計算の目的)」も説明できることになる。しかし、この必要十分性を導くためには、沢山の計算の方略を提案し、聴覚心理実験・生理実験によりこれらを検証することで、正しいものに絞り込まなければならないため、多くの時間を必要とする。

そこで、本論文では、「二つの音を分離する」という基本的な聴覚の機能に着目し、不良設定問題である信号分離問題を、二波形が加算されたものから個々の波形に分離抽出するという「二波形分離問題」に限定した。そして、計算理論を構築するための一歩として、「どのような制約条件を用いることで二波形分離問題を一意に解くことができるか」という戦略的な解法、つまり妥当と思われる聴覚の計算の方略の構築を試みた。

第２章では、音の分離抽出における聴覚の計算の方略を構築するための方法論を提案した。はじめに、本論文で取り扱う信号分離問題として二波形分離問題の枠組を示した。次に、本論文で取り扱う信号音の物理的表現を AM–FM 複合音で定義した。そして、Bregman によって提唱された四つの発見的規則を再考することで、二波形分離の対象となる分離抽出音を AM–FM 調波複合音と仮定した。また、二波形分離問題で利用する制約条件の思想

的な対応関係を示した。最後に、音の分離抽出に関する聴覚の計算の方略を構築するための方法論として、発展的構築法を提案した。

第３章では、二波形分離問題の理論的検討を行うことでその解法を提案した。これは、AM–FM 調波複合音を分離対象として取り扱える二波形分離問題を定式化し、観測された混合信号の瞬時振幅と瞬時出力位相から四つのパラメータ（二波形の瞬時振幅と瞬時入力位相）を一意に解けないことを示した。この結果から、本論文で扱う二波形分離問題が不良設定問題であることを明かにした。そこで、この不良設定問題を一意に解くために、Bregman によって提唱された四つの発見的規則を数理工学的な制約条件として定式化した。最後に、二波形分離問題の解法を忠実に実装し、AM–FM 調波複合音を分離抽出できることをシミュレーションで示した。

第４章では、AM–FM 調波複合音を利用して、第３章で提案した二波形分離問題の解法の十分性と有効性を検証した。はじめに、AM 単一成分音を利用して、二波形分離問題の解法による分離精度を評価することで、瞬時振幅に対する制約条件の十分性を実証した。次に、AM 調波複合音ならびに AM–FM 調波複合音を利用して、二波形分離問題の解法による分離精度を評価することで、瞬時位相ならびに基本周波数に対する制約条件の十分性を実証した。最後に、AM–FM 調波複合音を利用し、二波形分離問題の解法で利用する制約条件を順次省略した場合の分離精度を評価することで、制約条件の有効性を実証した。

第５章では、前章の検証結果から、二波形分離問の解法における制約条件を十分条件とした、音の分離抽出に関する聴覚の処理機能の「入力、出力、処理過程」を明らかにした。これは、検証された二波形分離問題の解法が、不良設定問題を一意に解くという問題を聴覚の情景解析の立場と数理工学の立場から統一的に議論して得られたものであることから、本解法を聴覚の計算の方略と解釈できた。以上の結果、音の分離抽出における聴覚の計算の方略とは、音の分離抽出という不良設定問題を区分線形問題と見なし、分離抽出したい信号の物理量の時間変化、つまり動きを拘束することで一意に解くことである、という結論を得た。この時、音の分離抽出における聴覚の処理機能の「入力、出力、処理過程」は

- 入力：分離抽出したい信号の物理量（瞬時振幅、瞬時位相、基本周波数）
- 出力：信号波形（あるいは瞬時振幅、瞬時位相、基本周波数の物理量）
- 処理過程：不良設定問題を区分線形問題と見なし、各物理量の時間変化を拘束しそれぞれがなめらかであるものから振幅包絡間の相関が最大になるように各物理量を取り出すこと。

である。

第６章では、音の分離抽出における聴覚の計算の方略を (1) 実音声（母音）を対象とした二波形分離問題、(2) 共変調マスキング解除を想定した二波形分離問題に展開することで、本計算の方略がこれらの問題の解法を導出できることを実証した。(1) の結果から、音の分離抽出における聴覚の計算の方略が音声認識のフロントエンドとしても応用できることを示した。(2) の結果からは、本計算の方略を展開することで、これまでに計算モデルが提案されていなかった CMR の計算モデルを実現できることを示した。

最後に、本論文では、計算の方略を導く際、制約条件の十分性と有効性しか議論していないが、制約条件の必要十分性を示すことで計算の方略を計算理論に発展させることができる。この必要十分性を導くためには、沢山の計算の方略を提案し、聴覚心理実験・生理実験によりこれらを検証することで正しいものに絞り込まなければならない。そのため、検証に多くの時間を必要とするが、本論文では、聴覚の計算理論を構築するための明確な方法論を提供できたと同時に、それに向けて確実に一歩前進したといえる。

7.2 今後の展望

以下に今後の展望を列挙する。

1. 制約条件の必要十分性の検証

   本論文では、音の分離抽出における聴覚の計算理論を構築するために、計算論的神経科学のアプローチで利用された方法論と同様、アルゴリズムの研究から計算理論の研究へと発展させることを試みた。従って、本研究では、聴覚系でどのような制約条件を設けることで目的の処理（不良設定問題を一意に解くこと）が可能なのかを、アルゴリズムより一段上のレベルで検討しなければならなかった。また、この制約条件を利用したアルゴリズムが計算理論から導かれたものになるためには、制約条件の必要十分性を示さなければならなかった。

   そこで、本論文では、計算理論を構築するための一歩として、「どのような制約条件を用いることで二波形分離問題を一意に解くことができるか」という戦略的な解法、つまり妥当と思われる聴覚の計算の方略の構築を試みた。具体的には、計算の方略の構築方法の一つとして発展的構築法を提案し、これを用いることで、二波形分離問題を一意に解くための制約条件をシミュレーションにて検証した。そして、最終的に、本解法で利用した制約条件を順番に省略した場合について、制約条件の有効性を検証した。この結果、本論文で導いた計算の方略では、制約条件に対する完全な必要条件を導いたわけではないが、少なくともこの条件があれば解けるという十分条件を導出

したことになる。また、音の分離抽出における計算の方略が明らかになったわけであるが、この方略はアルゴリズムの研究から計算理論の研究に発展させる過程で、アルゴリズムより一段上のレベルにあるが、まだ完全に計算理論のレベルにあるわけではない。従って、本論文で導いた計算の方略が完全に計算理論のレベルに発展するためには、本論文で利用した制約条件の必要十分条件を導かなければならない。この検証にはかなりの時間を要するものと予想されるが、二波形分離問題における様々な解法を提案し、その中から共通する制約を発見することで必要十分条件を検証できるものと考えられる。また、様々な聴覚心理実験や生理実験の結果を踏まえ、それを説明する計算モデルを実現し、検証することでも必要十分条件の議論が可能であると考えられる。

2. 計算論的な聴覚の情景解析の研究への方向性の提供

   第１章の研究背景でも述べたが、Cooke と Ellis は計算論的な聴覚の計算理論の研究から音のグルーピングに関する聴覚の計算理論の構築を試みている。しかし、取り扱う物理量は、振幅あるいはパワーのみであり、また Bregman の発展的規則の単なる実装に留まっているように見受けられる。本研究で得られた成果から、

   - 信号を厳密に分離するためには振幅の連続性だけではなく位相の連続性も重要であることを実証した。
   - 発見的規則を厳密に定式化し、十分性・有効性の検証を行った。
   - アルゴリズムの研究の域を出るために、計算理論の研究に発展するための構想を明確にした。

   という点で、彼らの研究とは異なる。これらは、計算理論の研究を発展させる意味で大きな強みでもある。従って、彼らの目指すグルーピングに関する計算理論の構築に向けても、本研究の成果は大きな影響を与えるものと考えられる。

3. 聴覚の情景解析の研究への方向性の提供

   Bregman によって提唱された四つの発見的規則は、第２章で説明したように様々な聴覚心理実験から得られたものであり、これがすべてであるとは言及されていない。また、これらの心理実験では、刺激音は音声のように実際的なものではなく、純音や複合音といった実験室レベルの人工音が取り扱われていた。本論文で明らかにされた計算の方略は、この四つの発見的規則を利用しているため、純音や調波複合音といった人工音はもちろん、AM–FM 調波複合音として解釈できる母音を対象に、二波形分

離問題を一意に解くことができる。しかし、子音は AM-FM 調波複合音で表現できないことから、この四つの発見的規則すべてで拘束されるとは考え難い。おそらく、子音の場合には四つの発見的規則のうちのいずれかだけを利用するか、あるいはまだ発見されていない規則を利用している可能性もある。そこで、本論文で明らかにされた計算の方略において、子音に対する二波形分離問題を解くために有効な制約条件を検証する。また、この問題を解くために別の制約条件が必要であれば、それを積極的に採り入れ、その理由を明らかにする。もし、これらの制約条件について心理学的・生理学的な意味を結びつけることができれば、これは聴覚の情景解析の研究に対する方向付けを示唆できる。また、これは、まだ発見されていない心理学的・生理学的知見を得るための実験の方向性も提供できる。

4. カクテルパーティー効果のモデル化に向けて

   本論文では、二波形分離問題における音の物理的表現と制約条件を明確にし、音を分離するためのアルゴリズムをボトムアップ的な構成で実装してきた。しかし、本論文で扱った問題は、二波形しか存在しないということと、目的音が意味のある波形であるという暗黙の仮定が存在した。そのため、カクテルパーティー効果のように、実際の環境の中からある特定の（誰の、何の）音を分離抽出するときには、文脈や文法といった音に関する知識ベース（トップダウン）の処理も必要になるものと考えられる。また、n 波形分離問題のような一般的な分離問題を解くことも必要になると考えられる。しかし、この問題については、n 波形分離問題を、望みの音とそれ以外の音（$n-1$ 個の他の音）とした二波形分離問題と解釈することで対応できるものと考えられる。将来的には、ボトムアップ処理である本モデルにトップダウン処理である知識ベース処理を組み合わせることで、実環境（複数音源が存在する）における実音声の分離抽出問題を解くことも可能であると考えられる。

5. 視覚の計算理論の研究と同じアプローチをとる研究への方向性の提供

   視覚の計算理論のアナロジーとして聴覚の計算理論を構築していく場合は、聴覚の心理学、生理学、情報科学の複合分野で統一的に議論していかなければならない。現段階で、このアプローチを取るには、聴覚の心理学的知見と生理学的知見が十分であるとはいえないが、マスキング現象に対する聴覚心理学および生理学の知見がかなりそろいつつある。本論文で提案した音の分離抽出における聴覚の計算の方略は、二つの音を分離するという場面を想定しているため、第６章に示したようにマスキング現象の問題に対応している。従って、共変調マスキング解除のように知られている様々なマスキング現象を二波形分離問題と解釈し、その問題を解くための制約条件を議論す

ることで、視覚の計算理論のアナロジーとして本研究を発展させていくことが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、終始、御指導ならびに御鞭撻を賜わりました北陸先端科学技術大学院大学　情報科学研究科助教授　赤木正人 博士に深甚なる感謝の意を表します。

筆者が青森職業訓練短期大学校在学中から、今日に至るまで終始ひとかたならぬ御指導と御教授を賜わったとともに、多大なる激励をいただきました青森職業能力開発短期大学校（旧名　青森職業訓練短期大学校）電子技術科助教授 高井秀悦 博士に深甚なる感謝の意を表します。

本研究を遂行するにあたり、熱心な御指導を賜わりました北陸先端科学技術大学院大学名誉教授（前北陸先端科学技術大学院大学教授、現（株）創研代表取締役） 飯島泰蔵 博士に深甚なる感謝の意を表します。

本研究を遂行するにあたり、日頃から熱心に御討論頂き、また有益なる御助言を賜わりました新潟大学助教授（前北陸先端科学技術大学院大学助手） 岩城護 博士に心より感謝致します。

本論文をまとめるにあたり、草稿の段階から貴重な御助言ならびに御指導を賜わりました北陸先端科学技術大学院大学教授 宮原誠博士に心より感謝致します。

本論文のまとめ、ならびに副テーマの遂行にあたり、貴重な御助言と御指導を賜わりました北陸先端科学技術大学院大学助教授 下平博 博士に心より感謝致します。

本研究を遂行するにあたり、日頃から熱心に御討論頂き、有益なる御助言を賜わりました和歌山大学教授 河原英紀 博士に心より感謝致します。

筆者が ATR 人間情報通信研究所に学外実習生として勤務したとき、そして本研究を遂行するにあたり、熱心な御指導ならびに御助言を賜わりました ATR 人間情報通信研究所主任研究員 入野俊夫 博士に心より感謝致します。

筆者が職業能力大学校在学中から今日に至るまで多大なる励ましをいただきました岩手大学助教授 西山清 博士並びに職業能力開発大学校教授 寺町康昌 博士に心より感謝致します。

また、日頃より多大なる議論と激励をいただきました北陸先端科学技術大学院大学の諸

先生方、飯島・赤木研究室時代の諸先輩方、並びに赤木研究室の諸氏に厚くお礼申し上げます。

なお、本研究の一部は、学術振興会特別研究員奨励金及び戦略的基礎研究推進事業（CREST）の援助によって行われたものです。深謝の意を表します。

最後に、私の研究生活を暖かく見守ってくれた両親、姉に心より感謝致します。

参考文献

[Abe and Ando, 1997] Abe, M and Ando, S. "Computational Auditory Scene Analysis Based on Loudness / Pitch / Timbre Decomposition," In Proc. IJCAI-97 Workshop on Computational Auditory Scene Analysis (CASA'97), pp. 47–54, Nagoya, Japan, August 1997.

[赤木, 1995] 赤木 正人, "カクテルパーティー効果とそのモデル化," 信学誌 vol. 78 no. 5 pp. 450-453, May 1995.

[赤木, 1998] 赤木 正人, "聴覚特性を考慮した波形分析," 音響誌 vol. 54 no. 8 pp. 575-581, August 1998.

[Bacon, 1989] Bacon, S. P. and Grantham, D. W. "Modulation masking: Effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. vol. 85, no. 6, June 1989.

[Berthommier *et al.*, 1995] Berthommier, M. and Meyer, G. "Source separation by a functional model of Amplitude Demodulation," In Proc. EUROSPEECH'95.4th, Madrid, Sept. 1995.

[Boll, 1979] Boll, S. F. "Suppression of Acoustic Noise in Speech using Spectral Subtraction," IEEE Trans. on Acoustic, Speech, and Signal Processing, Vol. ASSP-27, April, 1979.

[Bregman, 1990] Bregman, A.S. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press, Cambridge, Mass., 1990.

[Bregman, 1993] Bregman, A.S. "Auditory Scene Analysis: hearing in complex environments," in Thinking in Sounds, (Eds. S. McAdams and E. Bigand), pp. 10–36, Oxford University Press, New York, 1993.

[Brown, 1992] Brown, G.J. "Computational Auditory Scene Analysis: A Representational Approach," Ph. D. Thesis, University of Sheffield, 1992.

[Brown and Cooke, 1994] Brown, G.J. and Cooke, M.P. "Comutational auditory scene analysis," Computer Speech and Language, pp.297-336, 8, 1994.

[Brown and Hwang, 1992] Brown, R. G. and Hwang, P. Y. C. Introduction to Random Signals and Applied Kalman Filtering, second edition, Chapter 5–6, pp. 210–288, John Wiley and Sons, Inc., New York, 1992.

[Buus, 1985] Buus, S. "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. vol. 78, pp. 1958–1965, 1985.

[Cherry, 1953] Cherry, E.C. "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. vol 25, pp. 975–979, 1953.

[Cooke and Brown, 1993] Cooke, M.P. and Brown, G.J. "Computational auditory scene analysis : Exploiting principles of perceived continuity," Speech Communication, pp. 391-399, North Holland, 13, 1993.

[Cooke, 1991] Cooke, M. P. "Modeling Auditory Processing and Organization," Ph. D. Thesis, University of Sheffield, 1991 (Cambridge University Press, Cambridge, 1993).

[Cooke and Ellis, 1998] Cooke, M. P. and Ellis D.P.W. "THE AUDITORY ORGANIZATION OF SPEECH IN LISTENERS AND MACHINES," ICSI (International Computer Science Institute) Technical Report TR-98-016, June 1998.

[Chui, 1992] Chui, C.K. An Introduction to Wavelets, Academic Press, Boston, MA, 1992.

[Darwin *et al.*, 1992] Darwin, C. J., and Ciocca, V. "Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am. vol. 91, pp. 3381–3390, 1992.

[Darwin *et al.*, 1994] Darwin C. J., Ciocca, V., and Sandell, G.J. "Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic," J. Acoust. Soc. Am. vol. 95, no. 5 pp. 2631–2636, May, 1992.

[de Boor, 1978] de Boor, C. A Practical Guide to Spline. Springer-Verlag, New York, 1978.

[de Cheveigné, 1993] de Cheveigné, A. "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. 93, 3271–3290, 1993.

[de Cheveigné, 1997] de Cheveigné, A. "Concurrent vowel identification III: A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. 101, 2857–2865, 1997.

[Ellis, 1994] Ellis, D. P. W. "A Computer Implementation of Psychoacoustic Grouping Rules," Proc. 12th Int. Conf. on Pattern Recognition, 1994.

[Ellis, 1996] Ellis, D. P. W. "Prediction-driven computational auditory scene analysis," Ph. D. thesis, MIT Media Lab., 1996.

[Furui and Sondhi, 1991] Furui, S and Sondhi, M. M. Advanced in Speech Signal Processing, New York Marcel Dekker, Inc., 1991.

[古井, 1995] 古井. “音声認識,” 電子情報通信学会誌, vol.78 No.11 pp.1114-1118, 1995.

[古井, 1985] 古井. ディジタル音声処理, 東海大学出版会, 1985.

[Geman and Geman, 1984] Geman, S. and Geman, D. “Stochastic relaxation, Gibbs distributions and the bayesian restoration of images,” IEEE Trans. Pattern Anal. and Mach. Intell., PAMI-6, pp. 721-741, 1984.

[Gibson, 1979] Gibson, J.J. The Ecological Approach to Visual Perception, Houghton Mifflin Comany, Boston, 1979. (邦訳、古崎ら：生態学的視覚論, サイエンス社, 1985).

[Glasberg and Moore, 1990] Glasberg, B. R. and Moore, B. C. J., “Derivation of auditory filter shapes from notched-noise data,” Hear. Res., 47, pp. 103–138, 1990.

[Hall and Fernandes, 1984] Hall, J.W. and Fernandes, M.A. “The role of monaural frequency selectivity in binaural analysis,” J.Acoust. Soc. Am.76, pp.435-439, 1984.

[Hall and Grose, 1988] Hall, J.W. and Grose, J.H. “Comodulation masking release: Evidence for multiple cues,” J. Acoust. Soc. Am. vol. 84, pp. 1669–1675, 1988.

[Hansen and Nandkumar, 1995] Hansen, J. H. L. and Nandkumar, S. “Robust estimation of speech in noisy backgrounds based on aspects of the auditory process,” J. Acoust. Soc. Am. 97(6), June 1995.

[IJCAI-97 CASA, 1997] Computational Auditory Scene Analysis (CASA’97), 15th International Joint Conference on Artificial Interigence, Nagoya Congress Center, Nagoya, Japan, August 23-29, 1997.

[入野, 1991] 入野 俊夫, “聴覚末梢系表現からの信号再構成,” 音響学会聴覚研資, H-91-44, 1991.

[入野, Patterson, 1994] 入野 俊夫, Patterson, R.D. “音響事象検出・強調の計算理論,” 音響学会聴覚研資, H-94-64, Nov. 1994.

[入野, 1995a] 入野 俊夫, “聴覚末梢系の計算理論,” 信学技報, SP95-40, July 1995.

[入野, 1995b] 入野 俊夫, “最適聴覚フィルタの計算論的位置づけ,” 音響学会講演論文, 2-3-3, 1995.

[Irino and Patterson, 1997] Irino, T., Patterson, R.D. “A time-domain, level-dependent auditory filter: The gammachirp,” J. Acoust. Soc. Am. vol. 101 no. 1, 412–419, Jan. 1997.

[今井, 北村, 1978] 今井 聖, 北村 正, “対数振幅特性近似フィルタを用いた音声の分析合成系,” 信学論 Vol. J61-A No. 6, pp. 527–534, June 1978.

[Junqua and Haton, 1996] Junqua, J. C. and Haton, J. P. ROBUSTNESS IN AUTOMATIC SPEECH RECOGNITION, – fundamentals and applications –, Kluwer Academic Publishers, Boston, 1996

[柏野, 平原, 1997] 柏野 牧夫, 平原 達也, “文音声における同時複数話者の人数判断,” 音響学会春季講論, 2-8-12, March 1997.

[Kashino and Tanaka, 1993] Kashino, K. and Tanaka, T., “A sound source separation system with the ability of automatic tone modeling,” Proc. of Int. Computer Music Conference, pp. 248–255, 1993.

[柏野, 田中, 1994a] 柏野 邦夫、田中 英彦, “二つの周波数成分の分離知覚に関する工学的モデル,” 信学論 (A), vol. J77-A, no. 5, pp. 731-740, May 1994.

[柏野, 1994b] 柏野 邦夫, “計算機による聴覚の情景解析 – はじめの一歩 –,” 音響学会誌, vol. 50 no.12, pp. 1023-1028, Dec. 1994.

[柏野ら, 1996a] 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,” 信学論 (D-II), vol. J79-D-II, no. 11, pp. 1751-1761, Nov. 1996.

[柏野ら, 1996b] 柏野 邦夫, 木下 智義, 中臺 一博, 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,” 信学論 (D-II), vol. J79-D-II, no. 11, pp. 1762-1770, Nov. 1996.

[Katsuse *et al.*, 1997] Katsuse, I., Kawahara, H. and Aikawa, K. “Speech Segregation Based on Continuity of Spectral Shapes,” In Proc. IJCAI-97 Workshop on Computational Auditory Scene Analysis (CASA’97), pp. 39–45, Nagoya, Japan, August 1997.

[川人, 乾, 1990] 川人 光男, 乾 敏郎 視覚大脳皮質の計算理論, 信学論 (D-II) vol. J73-D-II, no. 8, pp. 1111-1121, August 1990.

[川人, 1996] 川人 光男, 脳の計算理論, 産業図書, 1996.

[河原, 1993] 河原 英紀, “聴覚の工学的表現,” 電子情報通信学会誌, vol. 11 pp. 1197-1202, 1993.

[河原, 1994a] 河原 英紀, “音声コミュニケーションにおける聴覚的情景解析,” 日本音響学会講演論文集, 2-7-13, 1994.

[河原, 1994b] 河原 英紀, “聴覚の計算理論の構築に向けて,” 音響学会聴覚研資, H-94-63, Nov. 1994.

[河原, 1991] 河原 英紀, “ウェーブレット解析の聴覚研究への応用,” 音響学会誌, vol. 47, no.6, pp. 424-429, 1991.

[Kawahara, 97] Kawahara, H. “STRAIGHT – TEMPO: A Universal Tool to Manipulate Linguistic and Para-Linguistic Speech Information,” In Proc. SMC-97, pp. 12-15, Orland, Florida, USA, Oct. 1997.

[Marr, 1982] Marr, D. Vision, Freeman, 1982. (邦訳, 乾, 安藤：ビジョン, 産業図書, 1987).

[McAulay and Quatieri, 1986] McAulary, R. J. and Quatieri, T. F. Low-Rate Speech Coding Based on the Sinusoidal Model, Advanced in Speech Signal Processing. Ed. Furui, S., pp. 165–208, Dekker, 1986.

[Moore, 1997] Moore, B.C.J. An Introduction to the PSYCHOLOGY OF HEARING, Forrth Edition, Academic Press, New York, 1997.

[Moore, 1992] Moore, B.C.J. "Comodulation Masking release and Modulation Discrimination Interface," in The Auditory Processing of Speech, from Sound to Words, (Edited by M. E. H. Schouten), pp. 167–183, Mouton de Gruyter, NewYork, 1992.

[中谷ら, 1993] 中谷 智広, 川端 豪, 奥乃 博, "計算論的アプローチによる音響ストリームの分離," 音響学会聴覚研資, H-93-83, 1993.

[Nakatani *et al.*, 1994] Nakatani, T. Okuno, H.G. and Kawabata, T. "Unified Architecture for Auditory Scene Analysis and Spoken Language Processing," In Proc. ICSLP'94, 24, 3, 1994.

[Nakatani *et al.*, 1995a] Nakatani, T. and Okuno, H. G., "A computational model of sound stream segregation with multi-agent paradigm,," In Proc. of ICASSP-95, Vol. 4, pp. 2671–2674, May 1995.

[Nakatani *et al.*, 1995b] Nakatani, T., Okuno, H. G., and Kawabata, T., "Residue-driven Architecture for Computational Auditory Scene Analysis," In Proc. of IJCAI-95, pp. 165–172, August 1995.

[中谷ら, 1995] 中谷 智広, 後藤 , 川端 豪, 奥乃 博, "調波構造と方向同定に基づく音響ストリーム分離," 音響学会秋季演論, 2-3-10, Sept. 1995.

[Natural Computation, 1988] "Natural Computation," W. Richards ed., MIT Press, 1988 (邦訳, 平原, 石川："ナチュラルコンピュテーション 2," パーソナルメディア, 1994).

[NATO ASI on Computational Hearing, 1998] NATO Advanced Study Institute on COMPUTATIONAL HEARING Edited by Steven Greenberg, Il Ciocco (Tuscany), 1-12 July 1998.

[西山, 中野, 1993] 西山 清, 中野 道雄, パソコンで解くカルマンフィルタ, 丸善, 1993.

[Okuno *et al.*, 1997] Okuno, G.H, Nakatani, T. Kawabata, T. "Challenge Problem for Computational Auditory Scene Analysis: Understanding Three Simultaneous Speeches," In Proc. IJCAI-97 Workshop on Computational Auditory Scene Analysis (CASA'97), pp. 61–68, Nagoya, Japan, August 1997.

[Papoulis, 1977] Papoulis, A. Signal Analysis. McGraw-Hill, New York, 1977.

[Papoulis, 1991] Papoulis, A. Probability, Random Variables, and Stochastic Process. Third Edition, McGraw-Hill, New York, 1991.

[Patterson and Holdsworth, 1991a] Patterson, R.D. and Holdsworth, J. "A Functional Model of Neural Activity Patterns and Auditory Images," Advances in speech, Hearing and Language Processing, Vol3. JAI Press, London, 1991.

[Patterson *et al.*, 1991b] Patterson, R.D. Robinson, K. Holdsworth, J. Mckown, D. Czhang and Allerhand, M. "Complex sounds and Auditory Images," 9th International Symposium on Hearing:Auditory physiology and perception, June 9-14, Carcans, France, 1991.

[Patterson *et al.*, 1995] Patterson, R.D., Allerhand, M., and Giguére, C. "Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform," J. Acoust. Soc. Am. vol. 98, pp. 1890-1894, 1995.

[Patterson and Moore, 1986] Patterson, R. D. and Moore, B. C. J. Auditory filters and excitation patterns as representations of frequency resolution. In Frequency Selectivity in Hearing (ed. B. C. J. Moore), Academic Press, London and New York, 1986.

[Poggio *et al.*, 1985] Poggio, T. Torre, V. and Koch, C. "Computational vision and regularization theory," Nature, 317, pp. 314–319, 1985.

[小特集「聴覚の情景解析」, 1994] "小特集「聴覚の情景解析」," 音響学会誌, vol. 50, no. 12, 1994.

[桜井, 1981] 桜井 明, スプライン補間入門, 東京電機大学出版局, 1981.

[Shamsunder, 1997] Shamsunder, S. and Giannakis, G. B. "Multichannel Blind Signal Separation and Recognition," IEEE Trans. on Speech and Audio Processing, vol. 5, No. 6, Nov. 1997.

[Takeda *et al.*, 1988] 武田 一哉, 匂坂 芳典, 片桐 滋, 阿部 匡信, 桑原 尚夫, 研究用日本語音声データベース利用解説書, ATR Technical Report TR-I-0028, 1988.

[Willen,*et al.*, 1992] Willen A. C. van den Brink, Tammo Houtgast, and Guido F. Smoorenburg. "Effectiveness of Comodulation Masking Release," in The Auditory Processing of Speech, from Sound to Words, (Eds. M. E. H. Schouten), pp. 167–183, Mouton de Gruyter, NewYork, 1992.

付録

A: 二波形分離問題の解の導出過程

式 (3.5) の両辺を２乗し、$\cos\theta_k(t)$ についてまとめると、

$$\cos\theta_k(t) = \frac{S_k(t)^2 - (A_k(t)^2 + B_k(t)^2)}{2A_k(t)B_k(t)} \tag{7.1}$$

を得る。この関係から、$\tan\theta_k(t)$ を求めるために、三角関数における角度 $\theta_k(t)$ と三角形の各辺の関係を考えると、各辺は

- 辺 a：$2A_k(t)B_k(t)$
- 辺 b：$Z_k(t) := S_k(t)^2 - (A_k(t)^2 + B_k(t)^2)$
- 辺 c：$\sqrt{(2A_k(t)B_k(t))^2 - Z_k(t)^2}$

であることがわかる。但し、$\cos\theta = b/a$、$\sin\theta = c/a$ である。つまり、

$$\cos\theta_k(t) = \frac{S_k(t)^2 - (A_k(t)^2 + B_k(t)^2)}{2A_k(t)B_k(t)} \tag{7.2}$$

$$\sin\theta_k(t) = \frac{\sqrt{(2A_k(t)B_k(t))^2 - Z_k(t)^2}}{2A_k(t)B_k(t)} \tag{7.3}$$

である。このことから、$\tan\theta_k(t)$ は

$$\tan\theta_k(t) = \frac{\sqrt{(2A_k(t)B_k(t))^2 - Z_k(t)^2}}{S_k(t)^2 - (A_k(t)^2 + B_k(t)^2)} \tag{7.4}$$

であることがわかる。ここで、上式の $\tan\theta_k(t) := Y_k(t)$ とおき、式 (3.22) の関係式にこれを代入することで、

$$\frac{S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))}{S_k(t)\cos(\phi_k(t) - \theta_{1k}(t)) - A_k(t)} = Y_k(t) \tag{7.5}$$

となる。但し、

$$Y_k(t) = \frac{\sqrt{(2A_k(t)B_k(t))^2 - Z_k(t)^2}}{Z_k(t)} \tag{7.6}$$

である。上式を整理すると、

$$\begin{aligned} Y_k(t)S_k(t)\cos(\phi_k(t)-\theta_{1k}(t))-Y_k(t)A_k(t) &= S_k(t)\sin(\phi_k(t)-\theta_{1k}(t)) \\ Y_k(t)S_k(t)\cos(\phi_k(t)-\theta_{1k}(t))-S_k(t)\sin(\phi_k(t)-\theta_{1k}(t)) &= Y_k(t)A_k(t) \\ R_k(t)\sin(\theta_{1k}(t)+\varphi_k(t) &= Y_k(t)A_k(t) \end{aligned}$$

を得る。従って、

$$\theta_{1k}(t)=\arcsin\left(\frac{A_k(t)Y_k(t)}{S_k(t)R_k(t))}\right)-\Phi_k(t) \tag{7.7}$$

ここで、$R_k(t)$ と$\Phi_k(t)$ はそれぞれ、

$$R_k(t)=\sqrt{2Y_k(t)^2+1} \tag{7.8}$$

と

$$\begin{aligned} \Phi_k(t) &= \arctan\left(\frac{Y_k(t)\sin\left(\phi_k(t)+\frac{\pi}{2}\right)+\sin\phi_k(t)}{\cos\phi_k(t)-Y_k(t)\cos\left(-\phi_k(t)+\frac{\pi}{2}\right)}\right) \\ &= \arctan\left(\frac{Y_k(t)\cos\phi_k(t)-\sin\phi_k(t)}{Y_k(t)\sin\phi_k(t)+\cos\phi_k(t)}\right) \end{aligned} \tag{7.9}$$

である。従って、これらをまとめると

$$\theta_{1k}(t)=\arcsin\left(\frac{A_k(t)Y_k(t)}{S_k(t)\sqrt{Y_k(t)^2+1}}\right)-\arctan\left(\frac{Y_k(t)\cos\phi_k(t)-\sin\phi_k(t)}{Y_k(t)\sin\phi_k(t)+\cos\phi_k(t)}\right)$$

を得る。

$\theta_{2k}(t)$ も同様の方法で導出できる。

■

B: 補題１の証明

はじめに、$\Psi_k(t)=\phi_k(t)-\theta_{1k}(t)$ とおく。式 (3.7) を整理すると、

$$A_k(t)=S_k(t)\cos\Psi_k(t)-S_k(t)\cot\theta_k(t)\sin\Psi_k(t). \tag{7.10}$$

を得る。上式の t について両辺を微分すると、

$$y'(t)+\frac{P'(t)}{P(t)}y(t)=\frac{Q'(t)}{P(t)}-\frac{C_{k,R}(t)}{P(t)}, \tag{7.11}$$

を得る。但し、$y(t) = \cot\theta_k(t),\ P(t) = S_k(t)\sin\Psi_k(t),\ Q(t) = S_k(t)\cos\Psi_k(t)$ である。この式は1次線形微分方程式であるから、一般解 $y(t)$ は

$$y(t) = \frac{1}{P(t)}\left(Q(t) - \int C_{k,R}(t)dt + C\right), \tag{7.12}$$

で得られる。但し、Cは未定係数である。従って、

$$\cot\theta_k(t) = \frac{S_k(t)\cos\Psi_k(t) - \int C_{k,R}(t)dt + C}{S_k(t)\sin\Psi_k(t)}, \tag{7.13}$$

から、次式を得る。

$$\theta_k(t) = \arctan\left(\frac{S_k(t)\sin(\phi_k(t) - \theta_{1k}(t))}{S_k(t)\cos(\phi_k(t) - \theta_{1k}(t) - C_k(t)}\right), \tag{7.14}$$

（証明終り）■

C: wavelet 変換の諸定義

はじめに、wavelet 分析合成系を設計するために必要な範囲で wavelet 変換の性質 [Chui, 1992] を以下にまとめる。

関数 $f(t)$ の wavelet 変換 $\tilde{f}(a,b)$ は、

$$\tilde{f}(a,b) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} f(t)\overline{\psi\left(\frac{t-b}{a}\right)}dt \tag{7.15}$$

なる積分変換で定義される。但し、a はスケールパラメータ、b はシフトパラメータであり、$\overline{\psi}$は ψの複素共役である。積分核は関数 $\psi(t)$ を a 倍のスケール変換と b だけシフトしたものとなっている。この関数 $\psi(t)$ の選択には数学的に大きな自由度をもつが、一般に許容条件 (admissibility condition)：

$$G_\psi := \int_{-\infty}^{\infty}\frac{|\hat{\psi}(\omega)|^2}{|\omega|}d\omega < \infty \tag{7.16}$$

を満たす正数 G_ψが存在するような二乗可積分関数とする。但し、$\hat{\psi}(\omega)$ は $\psi(t)$ の Fourier 変換である。ここで、$\hat{\psi}(\omega)$ が連続関数であるため、式 (7.16) の条件は、$\hat{\psi}(0) = 0$ あるいは$\int_{-\infty}^{\infty} f(t)dt = 0$ であることを課している。

$\psi(t)$ がこの許容条件を満たす時、$\psi(t)$ を基本 wavelet といい、次のような逆変換 (再構成) が存在する [Chui, 1992]。

$$f(t) = \frac{1}{G_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{f}(a,b)\psi\left(\frac{t-b}{a}\right)\frac{dadb}{a^2} \tag{7.17}$$

また、基本waveletが複素数値をとるとき、wavelet変換を振幅項$|\tilde{f}(a,b)|$と位相項$\arg(\tilde{f}(a,b))$で表すこともできる [河原, 1991]。

$$\tilde{f}(a,b) = |\tilde{f}(a,b)| \exp(j \arg(\tilde{f}(a,b))) \tag{7.18}$$

ここで、式 (7.15) と式 (7.17) は、連続 wavelet 変換対であるが、計算機に実装する際は、これに対応する離散 wavelet 変換を利用する。この変換対は、次式で表される。

$$\begin{aligned} \psi_{p,q}(t) &:= \alpha^{-p/2}\psi\left(\frac{t - q \cdot b_0}{\alpha^p}\right), \qquad b_0 = 1/f_s & (7.19)\\ \tilde{f}_{p,q} &:= \tilde{f}(\alpha^p, q/f_s) = \int_{-\infty}^{\infty} f(t)\overline{\psi}_{p,q}(t)dt & (7.20)\\ f(t) &= \frac{1}{G_{p,q}} \sum_p \sum_q \tilde{f}_{p,q}\psi_{p,q}(t) & (7.21) \end{aligned}$$

但し、p と qは整数値をとるパラメータである。

D: 補題２の証明

式 (7.18) の wavelet 変換は、式 (3.4) の分析フィルタの出力 $X_k(t)$ を複素数表現したものである。

$$\begin{aligned} X_k(t) &= S_k(t)e^{j(\omega_k t + \phi_k(t))} \\ &:= \tilde{f}(a,b), \qquad a = \alpha^{k-\frac{K}{2}}, b = t \end{aligned} \tag{7.22}$$

従って、両辺の絶対値を取ると

$$|X_k(t)| = S_k(t) = |\tilde{f}(\alpha^{k-\frac{K}{2}}, t)| \tag{7.23}$$

を得る。

（証明終り）■

E: 補題３の証明

式 (7.22) と式 (7.18) の位相項を比較すると

$$\omega_k t + \phi_k(t) = \arg(\tilde{f}(a,b)) \tag{7.24}$$

を得る。ここで、位相スペクトル $\arg(\tilde{f}(a,b))$ は、

$$\arg(\tilde{f}(a,b)) = \tan^{-1}\frac{Im\{\tilde{f}(a,b)\}}{Re\{\tilde{f}(a,b)\}} \tag{7.25}$$

と表されるため、$-\pi \leq \arg(\tilde{f}(a,b)) \leq \pi$ の区間ランプ関数となることがわかる。そこで、式 (7.24) の両辺を微分すると、

$$\omega_k + \frac{d\phi_k(t)}{dt} = \frac{\partial}{\partial t} \arg(\tilde{f}(\alpha^{k-\frac{K}{2}}, t))$$

となり、これを整理すると、

$$\frac{d\phi_k(t)}{dt} = \frac{\partial}{\partial t} \arg(\tilde{f}(\alpha^{k-\frac{K}{2}}, t)) - \omega_k$$

を得る。故に、出力位相 $\phi_k(t)$ は

$$\phi_k(t) = \int \left(\frac{d}{dt} \arg\left(\tilde{f}(\alpha^{k-\frac{K}{2}}, t)\right) - \omega_k \right) dt$$

となる。

（証明終り）■

本研究に関する研究業績

論文

[1] 鵜木 祐史, 赤木 正人, “雑音が付加された波形からの信号波形の一抽出法,” 信学論 (A), Vol. J80-A, No. 3, pp.444-453, March 1997. (Electronics and Communications in Japan, Part 3, Vol. 80, No. 11, pp. 1-11, 1997 Scripta Technica, Inc. Translated from IEICE).

[2] Masashi Unoki and Masato Akagi, “A Method of Signal Extraction from Noisy Signal based on Auditory Scene Analysis,” Speech Communication, Special Issue of Speech Communication on Computational Auditory Scene Analysis (in printing).

[3] Masashi Unoki and Masato Akagi, “A Computational Model of Co-modulation Masking Release,” NATO ASI series, IOS Press, Amsterdam, May 1999 (in preprinting).

[4] 鵜木 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,” 信学論 (A).（投稿中）

[5] 鵜木 祐史, 赤木 正人, “聴覚の情景解析に基づいた二波形分離モデルによる母音の分離抽出,” 信学論 (A).（投稿中）

国際会議

[1] Masashi Unoki and Masato Akagi, “A Method of Signal Extraction from Noisy Signal based on Auditory Scene Analysis,” In Proc. IJCAI-97 Workshop on Computational Auditory Scene Analysis (CASA’97), pp. 93-102, Nagoya, Japan, August 1997.

[2] Masashi Unoki and Masato Akagi, “A Method of Signal Extraction from Noisy Signal,” In Proc. EuroSpeech’97, Vol. 5, pp. 2587-2590, Rhodes, Greece, September 1997.

[3] Masashi Unoki and Masato Akagi, “A Computational Model of Co-modulation Masking Release,” In Proc. of NATO Advanced Study Institue on COMPUTATIONAL HEARING, pp. 129–134, Il Ciocco, Italy, July 1998.

[4] Masashi Unoki and Masato Akagi, “Signal Extraction from Noisy Signal based on Auditory Scene Analysis,” In Proc. of ICSLP’98, vol. 4, pp. 1515-1518, Dec. 1998.

研究会

[1] 鵜木 祐史, 赤木 正人, “雑音が付加された波形からの信号波形の抽出方法,” 音響学会聴覚研資, H-95-79, Nov. 1995.

[2] 鵜木 祐史, 赤木 正人, “共変調マスキング解除の計算機モデルに関する一考察,” 信学技報, SP96-37, July 1996.

[3] 鵜木 祐史, 赤木 正人, “帯域雑音中の AM 調波複合音の一抽出法,” 信学技報, SP96-123, March 1997.

[4] 鵜木 祐史, 赤木 正人, “基本周波数の時間変動を考慮した調波複合音の抽出法,” 信学技報, SP97-129, March 1998.

[5] 鵜木 祐史, 赤木 正人, “共変調マスキング解除の計算モデルの提案,” 音響学会聴覚研資, H-98-51, June 1998.

[6] 鵜木 祐史, 赤木 正人, “聴覚の情景解析に基づいた二波形分離モデルの提案,” 信学技報, SP98, March 1999.

口頭発表

[1] 鵜木 祐史, 赤木 正人, “帯域雑音に埋もれた信号音の抽出法,” 音響学会春季講論, 3-3-15, March 1996.

[2] 鵜木 祐史, 赤木 正人, “帯域雑音中の AM 調波複合音の一抽出法,” 音響学会春季講論, 2-8-7, March 1997.

[3] 鵜木 祐史, 赤木 正人, “共変調マスキング解除の計算モデルの高性能化,” 音響学会春季講論, 3-8-2, March 1997.

[4] 鵜木 祐史, 赤木 正人, “基本周波数の時間変動を考慮した調波複合音の抽出,” 音響学会春季講論, 2-8-16, March 1998.

[5] 鵜木 祐史, 赤木 正人, “聴覚の情景解析に基づいた雑音下の定常母音の分離抽出,” 音響学会秋季講論, 2-8-10, Sept. 1998.

[6] 鵜木 祐史, 赤木 正人, “聴覚の情景解析に基づいた二波形分離モデルの提案,” 音響学会春季講論, March. 1999.

その他の研究業績

論文

[1] Toshio Irino and Masashi Unoki, "An Analysis/Synthesis Auditory Filterbank Based on an IIR Implementation of the Gammachirp," The Journal of the Acoustical Society of Japan (E). (投稿中)

[2] Toshio Irino and Masashi Unoki, "A Time-Varying, Analysis/Synthesis Auditory filterbank based on an IIR Gammachirp filter," NATO ASI series, IOS Press, Amsterdam, May 1999 (in preprinting).

[3] 西山 清, 鵜木 祐史, "最大値を探索する人工ニューラルネットワーク," 信学論 (D-II), vol. J77-D-II, no. 7, pp. 1382-1385, July 1994.

国際会議

[1] Toshio Irino and Masashi Unoki, "A TIME-VARYING, ANALYSIS/SYNTHESIS AUDITORY FILTERBANK USING THE GAMMACHIRP," In Proc. ICASSP'98, vol. VI, pp. 3653-3656, Seattle, Washington, USA, May 1998.

[2] Toshio Irino and Masashi Unoki, "A Time-Varying, Analysis/Synthesis Auditory filterbank based on an IIR Gammachirp filter," In Proc. of NATO Advanced Study Institue on COMPUTATIONAL HEARING, pp. 205–210, Il Ciocco, Italy, July 1998.

研究会

[1] 入野 俊夫, 鵜木 祐史, "ガンマチャープフィルタとフィルタバンクの効率的な構成," 音響学会聴覚研資, H-97-69, Oct. 1997.

[2] 西山 清, 鵜木 祐史, "拡張 Hopfield 連想記憶モデルにおける冗長ニューロンの圧縮アルゴリズムの一般化," 信学技報, NC93-27, March 1993.

[3] 西山 清, 鵜木 祐史, "拡張 Hopfield 連想記憶モデル (I)," 信学技報, NC93-94, March 1994.

[4] 西山 清, 鵜木 祐史, “すべての記憶パターンが直交するように Hopfield 連想記憶モデルに付加された冗長ニューロンに付加された冗長ニューロンの圧縮可能性,” 信学技報, NC93-95, March 1994.

口頭発表

[1] 入野 俊夫, 鵜木 祐史, “IIR フィルタによるガンマチャープフィルタの実現,” 音響学会秋季講論, 1-3-16, Sept. 1997.

[2] 入野 俊夫, 鵜木 祐史, “ガンマチャープフィルタバンクによる時変系分析合成聴覚モデル,” 音響学会春季講論, 1-8-2, March 1998.

[3] 鵜木 祐史, 入野 俊夫, “ガンマチャープフィルタバンクにおける非対称性の制御方法,” 音響学会春季講論, 1-8-3, March 1998.