修 士 論 文

位相とスペクトルに着目した
聴覚の情景解析に関する基礎的研究

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

1996 年 2 月 15 日

Copyright © 2000 by Masashi Unoki

目 次

1 はじめに **1**

2 研究の背景 **4**

2.1 聴覚に関する生理学的・心理音響学的知見 4

2.1.1 位相同期 (phase-lock) 4

2.1.2 共変調マスキング解除 4

2.2 聴覚の情景解析 5

2.2.1 共通の立ち上がり/立ち下がりに関する規則 7

2.2.2 漸近的変化に関する規則 7

2.2.3 調波関係に関する規則 8

2.2.4 １つの音響事象に生じる変化に関する規則 8

2.3 聴覚の情景解析に基づく計算機モデルの研究 9

2.3.1 Brown と Cooke らの分凝モデル 9

2.3.2 Ellis の分凝モデル 12

2.3.3 中谷らの分凝モデル 14

2.3.4 柏野らの分凝モデル 15

2.4 本研究の特色 17

3 ２波形分離問題の定式化とモデルの概要 **18**

3.1 ２波形分離問題の定式化 18

3.2 モデルの概要 19

4 Gammatone filter を基底関数とした wavelet 分析合成系 **21**

4.1 Gammatone filter の特性 21

4.2 Gammatone filter を基底関数とした wavelet 変換 22

4.3 wavelet 分析合成系 24

5 物理パラメータの導出方法 **26**
5.1 振幅包絡 $S_k(t)$ の計算方法 26
5.2 出力位相 $\phi_k(t)$ の計算方法 27
5.3 入力位相 $\theta_k(t)$ の計算方法 28
5.4 未定係数の決定方法 30
5.5 2 波形分離アルゴリズム 33

6 2 波形分離のシミュレーション **36**
6.1 実験データ 36
6.2 分離結果 38
6.3 考察 44

7 人間の聴覚系に合わせたパラメータ設定 **50**
7.1 wavelet 分析合成系の再設計 50
7.2 実験データ 51
7.3 分離結果 52
7.4 考察 53

8 おわりに **55**

9 付録 **56**
9.1 付録 1 ：連続 wavelet 変換からの再構成 56
9.2 付録 2 ：入力位相 $\theta_k(t)$ に関する 1 階線形微分方程式の導出 57
9.3 付録 3 ：未定関数 $C(t)$ の性質 58
9.4 付録 4 ：未定係数 C_0 の諸性質 59
9.4.1 入力位相 $\theta_k(t)$ と出力位相 $\phi_k(t)$ に対する振幅包絡の関係 59
9.4.2 出力位相 $\phi_k(t)$ と未定係数 C_0 に対する入力位相と各振幅包絡の関係 60
9.4.3 未定係数 C_0 と振幅包絡 $A_k(t)$、$B_k(t)$ の関係 62

図 目 次

2.1 共変調マスキング解除の概略 6
2.2 共通の立ち上がり/立ち下がりに関する規則 7
2.3 漸近的変化に関する規則 8
2.4 調波関係に関する規則 9
2.5 １つの音響事象に生じる変化に関する規則 10
2.6 Brown と Cooke らの分凝モデルの構成 11
2.7 Ellis の分凝モデルの構成 13
2.8 中谷らの分凝モデルの構成 14
2.9 柏野らの分凝モデルの構成 16

3.1 Gammatone filter を基底関数とした wavelet 分析合成系の構成 19
3.2 モデルの概要 20

4.1 Gammatone filter のインパルス応答と周波数特性 (中心周波数：600[Hz]、$n = 4$,$b_f = 22.9945$) 22
4.2 wavelet 分析合成系の周波数特性 ($N = 128$) 25

5.1 未定係数 C_0 の範囲 32
5.2 隣接する聴覚フィルタの特性 33
5.3 ２波形分離アルゴリズム 35

6.1 音響信号：$f_1(t)$(上)、$f_{21}(t)$(中)、$f_{22}(t)$(下) 37
6.2 混合信号：$f_C(t)$(上) と $f_M(t)$(下) 37
6.3 混合信号：$f_{11}(t)$(上) と $f_{12}(t)$(下) 39
6.4 混合信号：$f_R(t)$(上) と $f_A(t)$(下) 39
6.5 振幅包絡と位相の時間微分：$\frac{dS_k(t)}{dt}$ と $\frac{d\phi_k(t)}{dt}$ 40

6.6 混合信号 $f_C(t)$ の場合：振幅包絡 $S_k(t)$(上)、出力位相 $\phi_k(t)$(中)、入力位相 $\theta_k(t)$(下) 41
6.7 混合信号 $f_C(t)$ の場合：分離された信号の振幅包絡 $A_k(t)$(上) と $B_k(t)$(下) . 42
6.8 混合信号 $f_C(t)$ の場合：再構成された信号 $\hat{f}_1(t)$(上) と $\hat{f}_2(t)$(下) 42
6.9 混合信号 $f_M(t)$ の場合：振幅包絡 $S_k(t)$(上)、出力位相 $\phi_k(t)$(中)、入力位相 $\theta_k(t)$(下) 43
6.10 混合信号 $f_M(t)$ の場合：分離された信号の振幅包絡 $A_k(t)$(上) と $B_k(t)$(下) 45
6.11 混合信号 $f_M(t)$ の場合：再構成された信号 $\hat{f}_1(t)$(上) と $\hat{f}_2(t)$(下) 45
6.12 混合信号 $f_R(t)$ の場合：振幅包絡 $S_k(t)$(上)、出力位相 $\phi_k(t)$(中)、入力位相 $\theta_k(t)$(下) 46
6.13 混合信号 $f_R(t)$ の場合：分離された信号の振幅包絡 $A_k(t)$(上) と $B_k(t)$(下) . 47
6.14 混合信号 $f_R(t)$ の場合：再構成された信号 $\hat{f}_1(t)$(上) と $\hat{f}_2(t)$(下) 47
6.15 混合信号 $f_A(t)$ の場合：振幅包絡 $S_k(t)$(上)、出力位相 $\phi_k(t)$(中)、入力位相 $\theta_k(t)$(下) 48
6.16 混合信号 $f_A(t)$ の場合：分離された信号の振幅包絡 $A_k(t)$(上) と $B_k(t)$(下) . 49
6.17 混合信号 $f_A(t)$ の場合：再構成された信号 $\hat{f}_1(t)$(上) と $\hat{f}_2(t)$(下) 49

7.1 周波数と ERB の関係 51
7.2 再設計した wavelet 分析合成系の周波数特性 52
7.3 帯域幅とマスキング解除の関係 53

9.1 各位相と振幅包絡 $A_k(t)$ の関係 59
9.2 $\phi_k(t)$–$A_k(t)$ 特性 (上) と $\theta_k(t)$–$A_k(t)$ 特性 (下) 60
9.3 各位相と振幅包絡 $A_k(t)$、$B_k(t)$ の関係 61
9.4 出力位相 $\phi_k(t)$ と未定係数 C_0 に対する入力位相 $\theta_k(t)$ の関係 61
9.5 振幅包絡 $A_k(t)$ の解析範囲 63

表 目 次

7.1 隣接する聴覚フィルタ数–$\hat{f}_1(t)$ の SN 比の関係 54

第 1 章

はじめに

高い柔軟性を持ち、雑音などの変動に強い音声認識 (機械認識) を実現するために、人間が音声を認識する仕組みを明らかにする研究が、聴覚の末梢から中枢に至るまで、神経生理モデルや音声知覚の観点から幅広く行なわれてきた [1]。

従来の音声認識手法は、(1) 雑音で汚れていないきれいな音声から、スペクトラムやケプストラムといった特徴量を分析し、(2) 抽出された特徴量を参照パターンとして、未知入力パターンとの間の距離測度をとり、未知パターンの分類を行なうことである。特に、(1) では、線形予測分析 (最ゆうスペクトル推定法) やケプストラム法といった特徴分析法 [2]、(2) では、DTW(Dynamic Time Warping)、HMM(Hidden Markov Model)、LVQ(Lerning Vector Quantization)、統計的言語モデルといった認識アルゴリズムが発展した [2, 3]。また、(1) と (2) を組み合わせた人工ニューラルネットワークといった技術も発展した [4]。これらの技術は、雑音による影響が少ないといった、かなり制約された状況下において、単独発話あるいは連続発話された音声の大語彙音声認識 (機械認識) を可能とした。しかし、雑音が付加されたような状況においては、未知パターンの特徴量の変動から、本来属するカテゴリーの参照パターンとの距離が大きくなり、誤認識の原因となるため、認識率の低下を招いていた。この問題点を解決するため、雑音が付加された音声から特徴抽出を行い参照パターンを作成する方法、音声のスペクトルを強調する方法、汎化能力の優れた距離測度を用いる方法、フィルタリング技術を利用して雑音を除去 (推定) する方法などが提案されているが、突発的な雑音にはとても弱い音声認識技術であった。

また、近年の聴覚生理研究の成果により、基底膜の有毛細胞や聴神経レベルの一部機能までを含んだ聴覚モデルが提案され、これを音声認識の前処理として積極的に取り入れようという動きもある [1]。この前処理は、聴覚フィルタを模擬した特徴抽出法の形態をとっており、これを用いた音声認識は耐雑音能力をもち、従来法よりも優れているという報告

もある [1, 9]。しかし、これもまた雑音が大きくなるに従って、認識率の低下は免れない。このように、約 40 年間に渡る音声認識の研究は、信号処理技術の観点と生理学的な観点から、様々な変遷を遂げてきたが、そこには雑音という強敵がいつでも潜んでいた。

一方、近年の聴覚心理研究の成果や生物の聴覚系に関する比較研究の成果により、聴覚は能動的な環境把握システムの一環として考えられ、聴覚の工学的表現や機能の見直しが進められている [5]。これは、McGurk 効果といったマルチモーダルな現象、耳からの音の放射の発見、共変調マスキング解除といった現象などから、最近の聴覚研究に大きな影響を与えている。また、2 つ以上のメッセージが混在していても、一方を選択的に聴取可能であるような “カクテルパーティー効果” という聴覚上の効果も研究されてきている。この効果がおこる原因は、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と考えられているが、まだはっきりとしたことがわかっていない [9]。しかし、近年、音を通じて環境をどのように把握するか？という観点から、聴覚の情景解析 (Auditory Scene Analysis)[6, 7] に関する研究結果が報告されつつあり、これらの知覚現象を説明する重要な鍵となっている [38]。この側面から、カクテルパーティー効果あるいは音源分離をモデル化することにより、聴取者にとって必要な音だけを選択し、他の音を除外するということが可能になる。従って、このモデル化を実現できれば、雑音で汚れていないきれいな音声だけを取り出すことが可能となり、これを前処理とすることで、雑音に強い音声認識への新たな側面を提供できる [9]。また、Marr[10] による視覚の計算理論のような、聴覚の計算理論の構築 [11, 12, 13] に向けても新たな視点を提供できる。

この聴覚の情景解析に基づく計算機モデルの研究として、いくつかの実装例が報告されている [25]。代表的なモデルとしては、Sheffield 大学の Brown と Cooke ら [15, 16, 17, 18] による音響イベントに着目した分凝モデル、MIT の Ellis[19] による心理音響学的グルーピング規則を取り入れた分凝モデル、NTT の中谷ら [20, 21] による聴覚の情景解析をマルチエージェントシステムによって実現した分凝の実装例がある。また、東大の柏野ら [23, 24] は、2 つの周波数成分の分離知覚に関して、スペクトログラム上の複数の特徴と分離知覚の生じる割合との定量的関係をモデル化している。しかし、いずれの計算機モデルも振幅スペクトル (パワー) を用いた分離を考えているため、2 つの信号が同じ周波数領域の成分を含むような場合、完全に分離できているとは言い難い。

我々は、同一周波数領域において完全に分離するためには、振幅スペクトルの他に位相も考慮しなければならないという立場に立ち [26, 27]、基本的な音源分離問題である 2 波形分離問題に取り組んでいる [28]。本論文では、この 2 波形分離問題の解法の 1 つとして、雑音が付加された波形から振幅スペクトルと位相の両方を考慮して信号波形を抽出する方法を提案する。また、本モデルを人間の聴覚特性に合わせてパラメータ設定した場

合、共変調マスキング解除の工学モデルになることを示す。

第 2 章では、聴覚に関する生理学的、心理音響学的知見と聴覚の情景解析について説明する。その後で、聴覚の情景解析に基づく計算機モデルについて、過去の研究の問題点を指摘し、本研究の特色を説明する。第 3 章では、本方法のモデルの概要を説明し、 2 波形分離問題の定式化を行なう。第 4 章では、本モデルの中心となる Gammatone filter を基底関数とした wavelet 分析合成系について、その設計方法と諸特性を示す。第 5 章では、定式化された物理パラメータの導出方法を示すと共に、これを用いた 2 波形分離アルゴリズムを説明する。第 6 章では、本手法の有効性を示すために、 2 波形分離のシミュレーション結果をまとめ、第 7 章では、分析合成系を人間の聴覚特性に合わせたパラメータ設定を行なうことにより、共変調マスキング解除の工学的モデル化が可能であることを示す。最後に、第 8 章で結論を述べる。

第 2 章

研究の背景

この章では、本研究の土台となる聴覚に関する生理学的、心理音響学的知見を述べる。また、Bregman によって提唱された “聴覚の情景解析 (Auditory Scene Analysis)” と発見的規則を説明し、これに基づくいくつかの計算機モデルを紹介するとともに問題点を列挙する。最後に、本研究の特色を述べる。

2.1 聴覚に関する生理学的・心理音響学的知見

2.1.1 位相同期 (phase-lock)

聴覚の神経発火は刺激音の波形に同期する傾向があり、これを位相同期 (phase-lock) という。このとき神経パルスの時間間隔は、音刺激の基本周期またはその整数倍となる。聴覚神経発火の位相同期性は、聴覚モデルによく取り入れられる特徴である。

この知見は、位相変化のずれが生じている (周期にずれが生じている) とき、 2 波形が存在していること示すことから、 2 波形分離の重要な手がかりとなる [26]。本研究では、この位相のずれを出力位相と定義し、これを 2 波形分離の手がかりとしている。詳細については、第 5 章で述べることにする。

2.1.2 共変調マスキング解除

2 つの音がほぼ同時期に存在するとき、一方の音によって最小可聴域が上昇し、他の信号音をマスクするような現象をマスキングとよんでいる。また、この最小可聴域の上昇をマスキング量といい、マスクする音をマスカーとよんでいる。このマスキングに関する聴覚心理実験は、Fletcher や Patterson、Moore らをはじめ、数多く行なわれているが、こ

れらは聴覚フィルタの帯域幅 (臨界帯域) を測定するために、マスカーの時間的変動を無視して行なわれていた。しかし、人間は単一の聴覚フィルタを使って音を聞くのではなく、フィルタ間の比較を行なうことを示す結果もある [31]。Hall らは、初めてフィルタ間の比較によって変動する雑音にマスクされた正弦波信号の検知が容易になる可能性を示した [32]。この実験では、マスカーの中心周波数を 1[kH] とし、 2 種類のマスカー：

- 振幅が不規則で、かつ異なる周波数領域で独立に変動する帯域雑音 (ランダム帯域雑音)
- 振幅が不規則で、かつゆっくりとした速度で変調した帯域雑音 (AM 帯域雑音)

を用意した。実験は、スペクトルレベルを一定にしたマスカーの帯域幅を変動して、1[kHz]、400[ms] の正弦波信号の閾値を測定した。この結果、図 2.1のように、マスカーの帯域幅が狭い ($\leq$ 130[Hz]) ところでは、ランダム雑音 (R) と変調雑音 (M) のいずれでも信号の閾値が増加したのに対し、帯域幅が広い ($\geq$ 130[Hz]) ところでは、ランダム雑音の場合、信号の閾値は変化せず、変調雑音の場合、信号の閾値が減少した。このマスキング量の減少は、共変調マスキング解除 (Co-modulation masking release) と呼ばれ、雑音の振幅変調を異なる周波数領域間で一致させることを共変調ということに由来している [32]。

この知見は、聴覚フィルタ間の振幅包絡を比較するよい理由となっており、第 5 章で未定係数 C_0 を決定するときの手がかりとなっている。また、上記のような実験データに対し、第 6 章では、 2 波形分離が可能であることを示し、第 7 章では、本研究のモデルが共変調マスキング解除の工学的なモデルになることを示している。

2.2 聴覚の情景解析

Bregman は、“Auditory Scene Analysis”[6] で多くの聴覚心理実験の結果と考察をまとめ、聴覚の情景解析 (Auditory Scene Analysis) を提案している。これは、今までに個別に研究されてきた様々な聴覚現象を、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。

聴覚は、情景解析の問題を容易に解いているように感じるかもしれない。しかし、我々が日常経験する複雑な (様々の音源からの直接音や反射音、残響、騒音といった) 環境において、周囲に何があり、何が起きているかを把握することは、音源や環境に関する制約がない限り、一意的な解の出ない困難な問題である [8]。この困難な問題を解くために、Bregman は、聴覚が利用している制約条件のいくつかを音響事象 (Acoustic events) に関

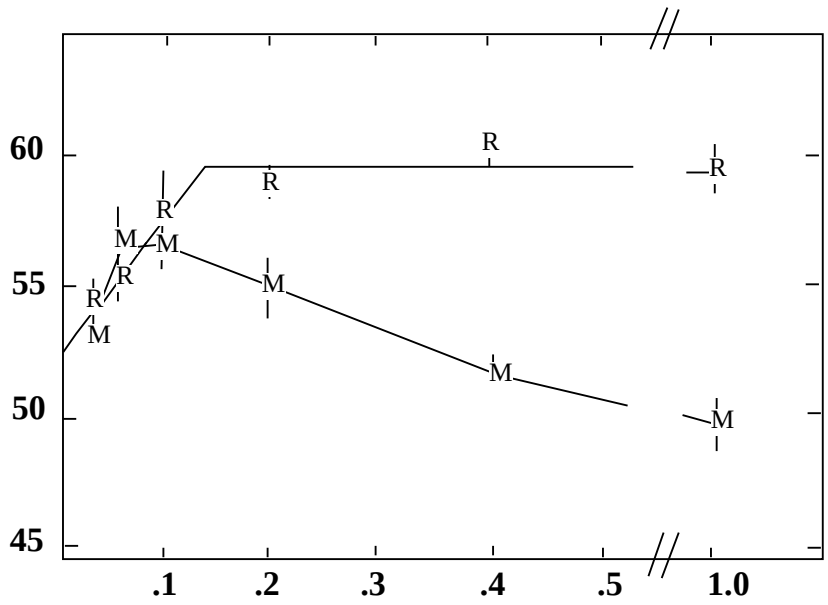
60
55
50
45
.1
.2
.3
.4
.5
1.0
R
R
R
R
R
M
M
R
R
M
M
M
M

図 2.1: 共変調マスキング解除の概略

係する 4 つの心理学的な発見的規則として述べている [7]。この発見的規則は、次の 4 つでまとめられている。

1. 共通の立ち上がり/立ち下がりに関する規則
2. 漸近的変化に関する規則
3. 調波関係に関する規則
4. 1 つの音響事象に生じる変化に関する規則

Bregman の説明によれば、聴覚は、この規則を用いて、入ってきた音のかたまりを複数の音のストリームに分け、外界の解釈を試みることになる。聴覚的誘導や音韻修復等、錯覚と考えられている現象は、規則 1 と 2 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 [8]。また、共変調マスキング解除も、規則 4 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。

本研究では、規則 1 と規則 4 を用いることで、 2 波形分離問題を解いている。詳細については、第 5 章で述べることにする。尚、上記の規則は、Bregman も指摘するように、これですべてであるとはいえず、現時点でこの 4 つの規則がわかっているにすぎない。

次に、これら 4 つの発見的規則について具体例をあげて説明する。

2.2.1　共通の立ち上がり/立ち下がりに関する規則

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” ということを示している。言い換えると、“ある 1 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” ことを意味している。

例えば、図 2.2に示すように、音源 A で生じた音は共通の立ち上がりと立ち下がりになるのに対し、音源 B で生じた音はこれとは共通な立ち上がり/立ち下がりにならない。

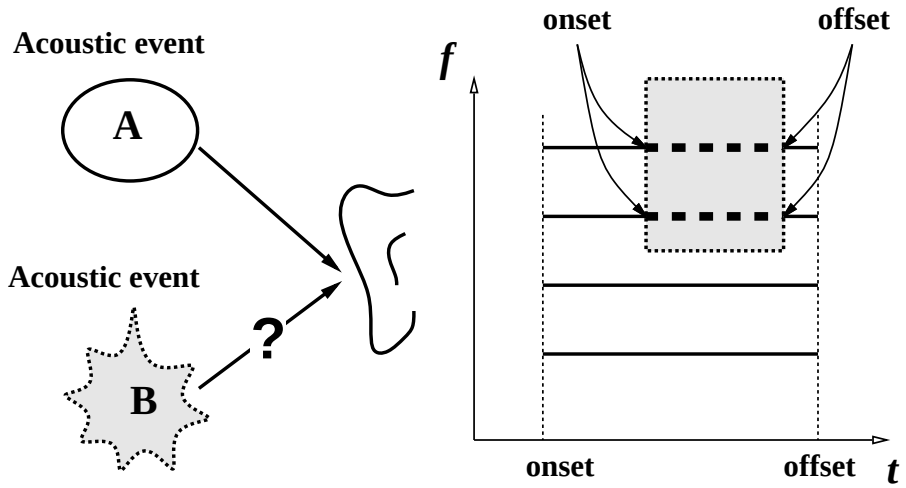

図 2.2: 共通の立ち上がり/立ち下がりに関する規則

2.2.2　漸近的変化に関する規則

この発見的規則は、漸近的な変化：

1. 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。

2. 同じ音源からくる一連の音は、その性質上ゆっくりと変化する傾向がある。

を示している。これは、同じ音源からくる音は急激に変化せず、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ている。

例えば、図 2.3に示すように、2 つの音源で生じた音が単耳に到来した場合を想定する。音源 A と音源 B で生じる音は、それぞれ漸近的変化の規則に従っているが、これらが混

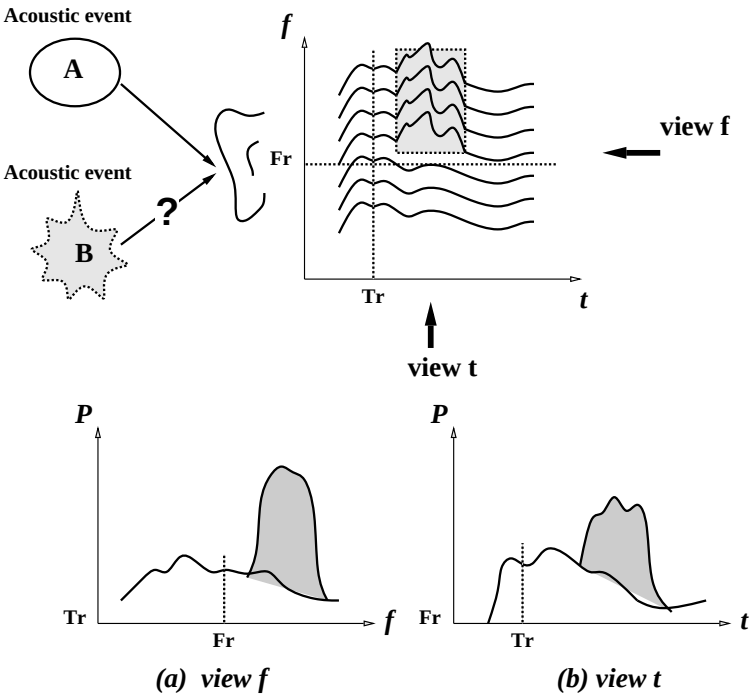

図 2.3: 漸近的変化に関する規則

合された時点で急激な変化を生じてしまう。図 $2.3(\mathrm{a})$ は、$t = T_r$ のときの周波数とパワーの関係を、図 $2.3(\mathrm{b})$ は、$f = F_r$ のときの時間とパワーの関係を表している。

2.2.3 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” ということを示している。これは、定常母音といった調波音に関する規則であり、調波関係を示している。

例えば、図 2.4に示すように、音源 A から調波音が、音源 B から帯域雑音が到来した場合を想定する。このとき、基本周波数を抽出し、この周波数の整数倍の関係にある周波数成分だけを抽出すれば、音源 A で生じる調波音を抽出できる。

2.2.4 1 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、“1 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” ということを示し、共変調マスキング解除のよい説明になって

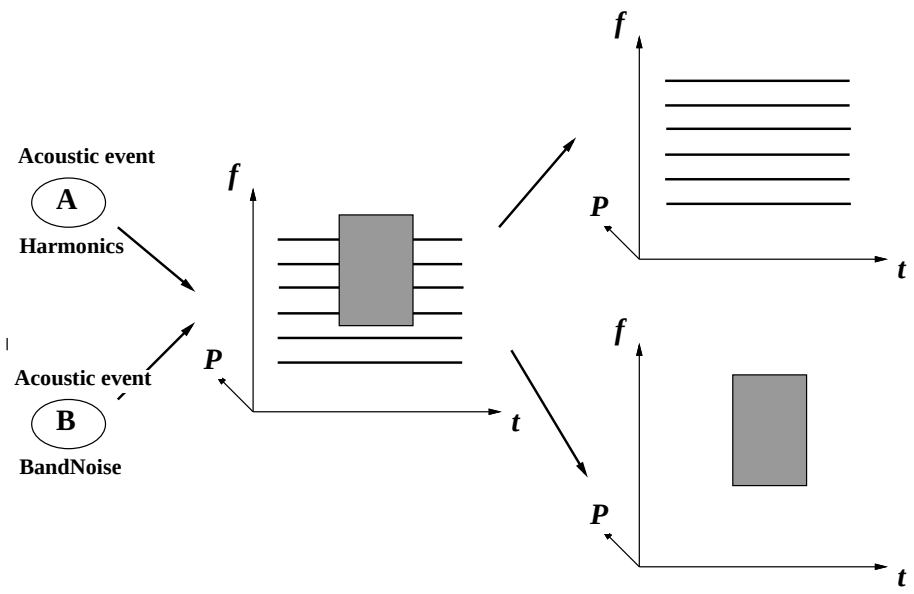

図 2.4: 調波関係に関する規則

いる。

例えば、図 2.5に示すように音源 A から純音が、音源 B から振幅変調された帯域雑音が到来した状況を想定する。本来、音源 B からランダム帯域雑音が到来した場合であれば、マスキング現象が起こり、このマスキング量は帯域幅の増加と共に増加し、ある帯域幅 (臨界帯域) を越えた時点で一定になる。しかし、振幅変調された帯域雑音の場合、帯域幅の増加と共にマスキング量が減少するという現象が起こった。これは、雑音の周波数成分間に高い相関があるため (あるいは一致するため)、これを手がかりにマスキングを解除しているという理由から共変調マスキング解除と呼ばれている。

2.3 聴覚の情景解析に基づく計算機モデルの研究

2.3.1 Brown と Cooke らの分凝モデル

Brown と Cooke らによる分凝モデル [15, 16, 17, 18] は、聴覚の情景解析に基づいた計算機モデルとして、最も先に提案されたモデルである。この分凝モデルの構成を図 2.6に示す。

このモデルは、聴覚末梢系 (periphery)、聴覚マップ表現 (auditory map representation)、音源分離 (source segregation) という 3 つの処理過程で構成される。はじめに、混合信号

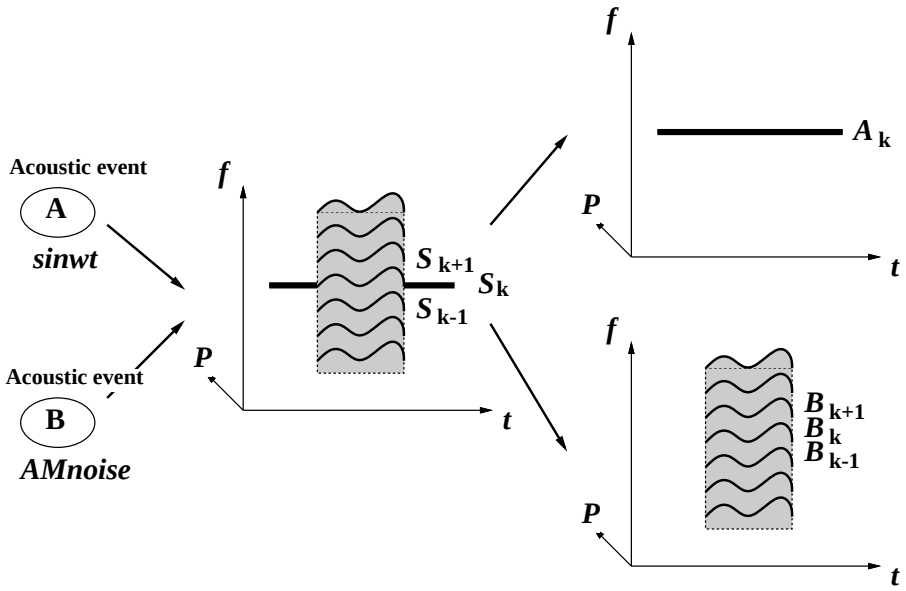

図 2.5: 1 つの音響事象に生じる変化に関する規則

は、Gammatone filtebank と神経発火モデルから成る聴覚末梢系により、computational map と呼ばれる抽象表現がとられる。これは、(1) 神経発火率 (fire rate)、(2) 周波数遷移 (frequency transition)、(3) 立ち上がり (onset)、(4) 立ち下がり (offset)、(5) 自己相関 (autocorrelation)、(6) フィルタ間の相関 (cross-correlation) といった map 表現である。他に、強度、強度差、スペクトル形状、同期性といったものもある [17]。次に、これらの抽象表現をもとに、聴覚要素 (auditory elements) が形成される。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。

このモデルの最大の特徴は、

- 分離に用いられる特徴はすべて生理学的あるいは心理学的知見に基づいていること。
- 音響事象 (acoustic event) と呼ばれる特徴量を用いていること。
- このモデルは、あらゆる雑音を音声から減少させることが可能であり、この減少量を定量的に示したこと。

であり、実際の分凝例として、音声 + サイレンなどの分離結果が示されている [17]。

しかし、

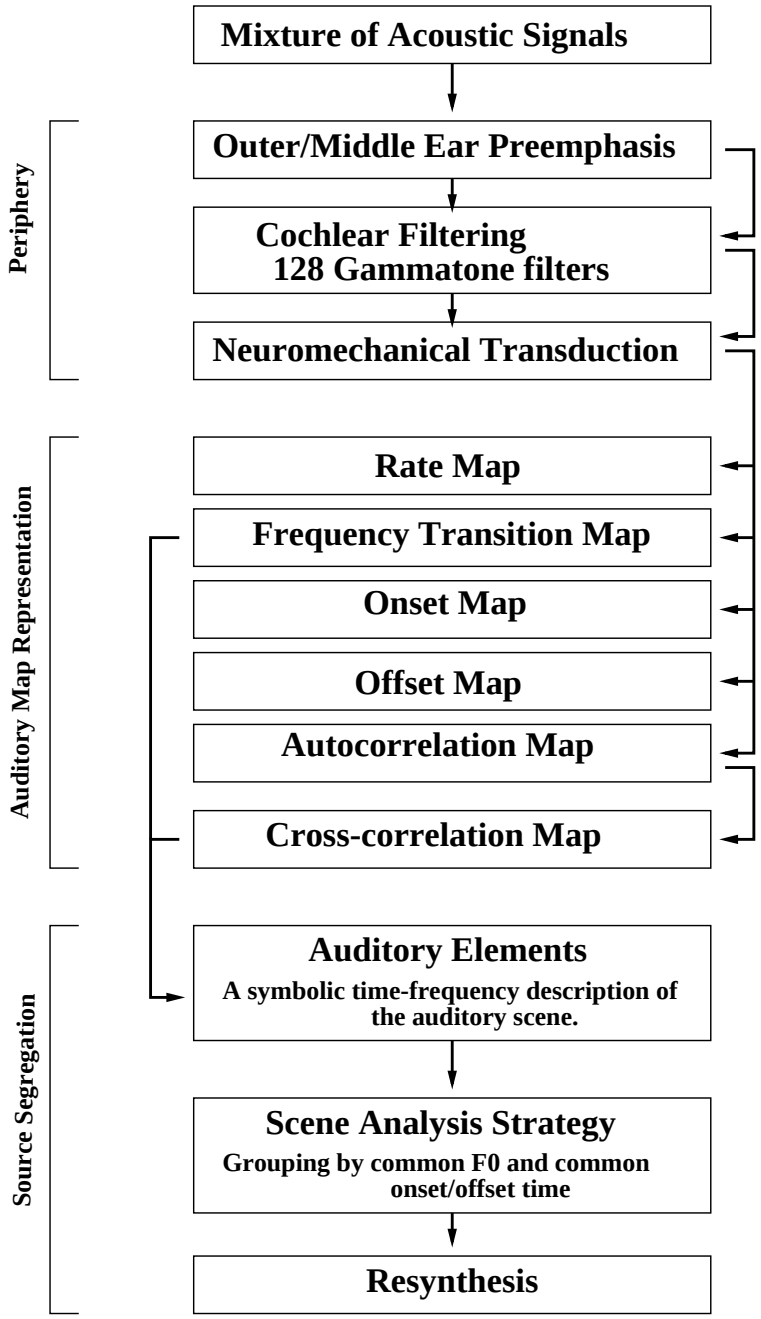

図 2.6: Brown と Cooke らの分凝モデルの構成

- 振幅スペクトルを用いて分離しているため、同一周波数領域 (あるいは同一フィルタ内) において完全に分離できているとはいえない。
- 特徴の評価における評価値の意味付けがあいまいである。

といった問題点がある。

2.3.2 Ellis の分凝モデル

Ellis の分凝モデル [19] は、人間の聴覚系で行なわれ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を実現しようとしたものであり、心理学的知見を重視したモデルである。この分凝モデルの概要を図 2.7に示す。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 (primary grouping)、(4) 2 次グルーピング (secondary grouping)、(5) 高次処理で構成されている。

はじめに、音響事象に対し、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る。次に、初期グルーピングにより、部分的なオブジェクト (part objects) を生成する。この初期グルーピングは、調波性、共通な立ち上がり、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。次に、2 次グルーピングにより、部分オブジェクトから基本オブジェクト (basic objects) 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 (pruning)、相関 (correlation)、接近 (closure) といった規則で構成されている。最後に、高次処理により、処理に基づいた知識などを得ている。

このモデルの特徴は、

- 心理学的知見を重視したモデルであること。
- グルーピングを 2 段階に分離し、初期グルーピングで考えられる候補をすべてを 2 次グルーピングで適切なグループに絞り込んでいること。

であり、分凝例として、“缶が硬い表面に落ちた音とソロクラリネットのメロディーが混ざった音” の分離結果を示している。しかし、このモデルには、

- 振幅スペクトルを用いて分離しているため、同一周波数領域において完全に分離できているとはいえない。

といった問題点がある。

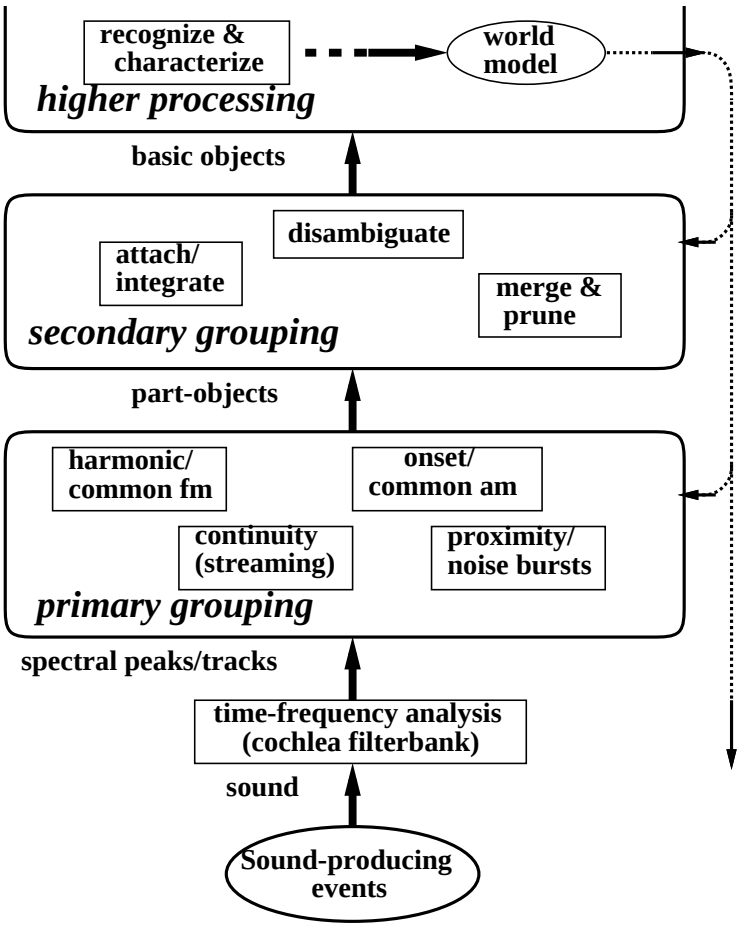

図 2.7: Ellis の分凝モデルの構成

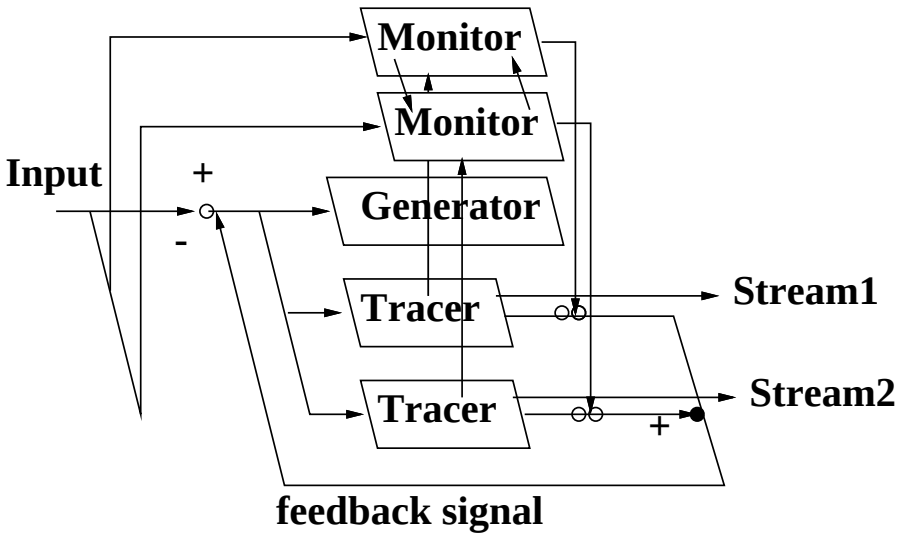

図 2.8: 中谷らの分凝モデルの構成

2.3.3 中谷らの分凝モデル

中谷ら [20, 21] によるモデルは、音情景解析 (auditory scene analysis) をマルチエージェントシステムにより実現したものであり、音の調波構造に注目した音響ストリーム分離エージェントを構築している。この分凝モデルの構成を図 2.8に示す。彼らは、音響ストリーム分離が実環境下で特定の音に注目して処理を行なう音環境理解を実現するための一部品であると位置付けている [20, 21]。ここで、入力 (input) として音の波形が逐次的に与えられるとき、音響ストリームを分離するための課題：

1. 新しい調波構造の発見と追跡エージェントの生成
2. 調波構造の追跡
3. ストリーム間の干渉の低減
4. 調波構造の終了の検知

を設定している。この課題を解くために、基本システムは 1 つの生成エージェント (Generator) と動的に生成され消滅する複数の追跡エージェント (Tracer) から成る。特に、課題 (1) を生成エージェントが、課題 (2)、(4) を追跡エージェントが担当し、課題 (3) をフィードバック信号が担当している。ここで、基本システムは時間的にも空間的にもローカルな連続性を利用して、ストリームを追跡するため、自然に消滅せず、いつまでも残り

続けるような冗長な追跡エージェントを生成するという問題があった。彼らは、基本システムに、モニターエージェント (Monitor) を加えることで、この問題点を解決している。これにより、

- 冗長な追跡エージェントを発見して除去する。
- 自己発振を検知して、ストリームの整形をする。

といった課題が拡張されている。これは拡張システムと呼ばれている。

このモデルの特徴は、

- マルチエージェントシステムの創発性は、システムが処理したい音を能動的に処理するための枠組として有用であること。
- エージェントの種類を増やすことにより、システムの機能が拡張できるという開放系が容易に構成できること。
- エージェントが競合する中から様々な行動が取られるという創発性が容易に実現できること。

であり、分凝例として、男性の声と合成音から成る混合音の分離や、男性と女性の声の混合音の分離結果が示されている。また、 2 チャンネル入力から抽出される方向情報と音源方向同定による手がかりを用いて、更に分離性能 (スペクトル歪み、基本周波数誤差の減少) を向上できることも報告されている [22]。

しかし、このモデルにも、

- 振幅スペクトルを用いて分離しているため、同一周波数領域において完全に分離できるとはいえない 。

という問題点がある。

2.3.4 柏野らの分凝モデル

柏野ら [23, 24] は、知覚的音源分離システムを実現するための基礎として、 2 つの周波数成分の分離知覚に関するモデルを提案している。

ここでは、音源分離問題を

1. 混合音を分析して得たサウンドスペクトログラム上で、周波数成分を抽出する問題 (周波数成分の抽出問題)

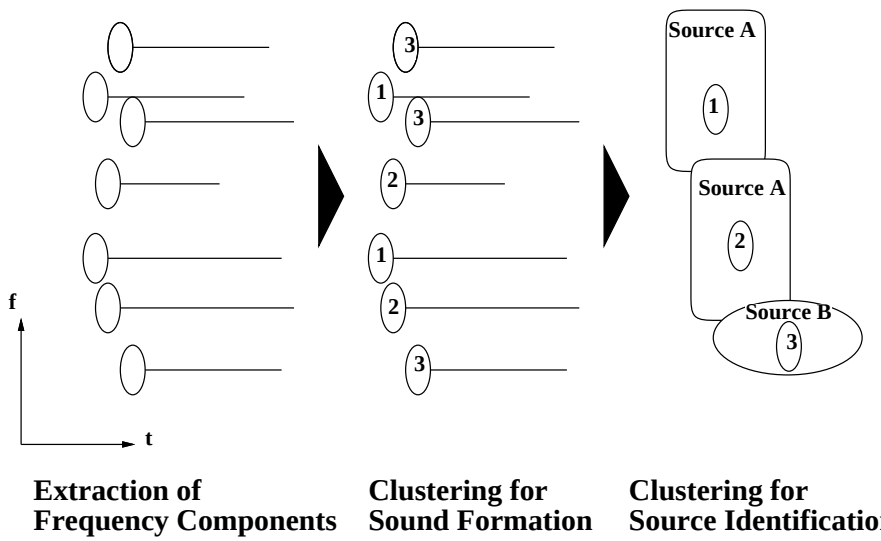

図 2.9: 柏野らの分凝モデルの構成

2. 得られた周波数成分を、重複を許してある個数のクラスタに分類する問題 (周波数成分のクラスタリング問題)

と定式化している。また、問題 (2) は、更に、 2 段階処理に分けて考えられている。この処理の概要を図 2.9に示す。

はじめに、抽出された周波数成分に対し、人間が 1 つの音として知覚しやすい周波数成分同士をまとめるような単音形成クラスタリング (clustering for sound formation) を行なう。これにより、単音クラスタが生成される。次に、単音クラスタに対し、同じ音源に由来すると考えられるもの同士をまとめる音源同定クラスタリング (clustering for source identification) を行なう。これにより、音源クラスタが得られ、同一の音源クラスタに属する単音クラスタを時間の順に並べることにより、ある音源に由来する音響的情報を抽出することができる。

このモデルの特徴は、

- 2 つの周波数成分の分離知覚に関して、複数の特徴と分離知覚の生じる割合との関係を定量的にモデル化していること。

であるが、

- 振幅スペクトルを利用していることと、重複周波数成分の分離を単音クラスタの重複を許すことで行なっているため、同一周波数領域において完全に分離できているとはいえない。

といった問題点がある。

2.4 本研究の特色

第 2.3 節で述べたように、いずれの分凝モデルも振幅スペクトルを用いて分離を行なっているため、 2 つの信号が同じ周波数領域を含むような場合、完全に分離できているとは言い難い。

本研究では、同一周波数領域において完全に分離するためには、振幅スペクトル (パワー) の他に位相も考慮しなければならないという立場に立ち、基本的な音源分離問題の 1 つである 2 波形分離問題に取り組んでいる [28]。

また、本研究の主な特色は、

- 聴覚フィルタ群は、基底膜の特性を模擬している Gammatone filter と、これを基底とする wavelet 分析系で構成されること [27, 28]。
- 心理学的、心理音響学的知見を取り入れていること。
- 分離規則とグルーピング規則は、Bregman が提唱した 4 つの発見的規則のうち、
  1. 漸近的変化に関する発見的規則
  2. 1 つの音響事象で生じる変化に関する発見的規則

  を物理的制約条件として記述していること。

である。

第 3 章

2 波形分離問題の定式化とモデルの概要

3.1 2 波形分離問題の定式化

本研究では、ある 2 つの音響信号が合成された状況から、それぞれの音響信号に分離・再合成することを、 2 波形分離問題と定義する。この 2 波形分離問題は以下のように定式化される。

はじめに、ある 2 つの音響信号波形 $f_1(t)$ と $f_2(t)$ が信号

$$f(t) = f_1(t) + f_2(t) \tag{3.1}$$

に合成された状況を想定する。これらは、図 3.1に示した N 個の聴覚フィルタで構成される $\mathrm{wavelet}$ 分析系により周波数分解される。次に、$f_1(t)$ と $f_2(t)$ がそれぞれ k 番目の分析フィルタにより

$$f_1(t) \quad : \quad A_k(t)\sin(\omega_k t) \tag{3.2}$$

$$f_2(t) \quad : \quad B_k(t)\sin(\omega_k t + \theta_k(t)) \tag{3.3}$$

に周波数分解されたものとする。但し、ω_k は分析フィルタの中心角周波数、$\theta_k(t)$ は $f_2(t)$ のもつ $f_1(t)$ に対応した入力位相である。このとき、分析フィルタの出力 $X_k(t)$ は、

$$\begin{aligned} X_k(t) &= A_k(t)\sin\omega_k t + B_k(t)\sin(\omega_k t + \theta_k(t)) \\ &= S_k(t)\sin(\omega_k t + \phi_k(t)) \end{aligned} \tag{3.4}$$

と表される。また、振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ は、それぞれ

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

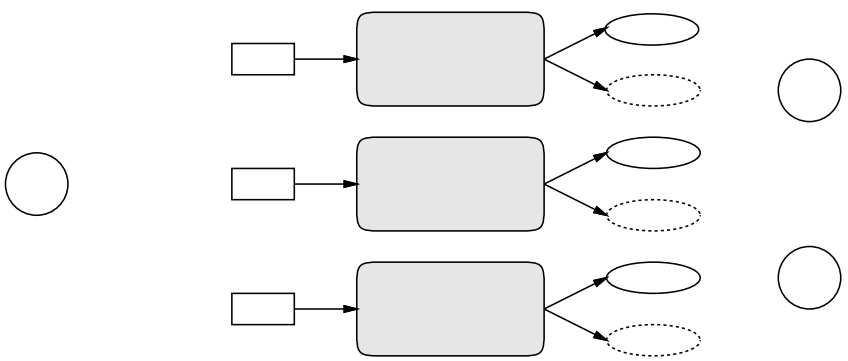

図 3.1: Gammatone filter を基底関数とした wavelet 分析合成系の構成

$$\phi_k(t) = \tan^{-1}\left(\frac{B_k(t)\sin(\theta_k(t))}{A_k(t) + B_k(t)\cos\theta_k(t)}\right) \tag{3.6}$$

となる。ここで、振幅包絡 $S_k(t)$ と周波数 ω_k が観測可能であることから、出力位相 $\phi_k(t)$ と入力位相 $\theta_k(t)$ がわかれば、 2 波形の振幅情報 $A_k(t), B_k(t)$ を

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

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

のように解析的に解くことができる。この処理過程は、図 3.1の分離部で行なわる。最後に、すべての分析フィルタ $(X_k(t), 1 \leq k \leq N)$ について、振幅包絡 $A_k(t)$ と $B_k(t)$ に分離し、それぞれを同じ信号成分のものとして wavelet 合成系により合成 (グルーピング) することで、$f_1(t)$ と $f_2(t)$ を再構成できる。但し、再合成された信号を、それぞれ $\hat{f}_1(t)$、$\hat{f}_2(t)$ とする。

3.2 モデルの概要

本研究のモデルの概要を図 3.2に示す。はじめに、混合信号 $f(t)$ が聴覚フィルタ群 (Auditory filterbank) により、振幅包絡 $S_k(t)$、出力位相 $\phi_k(t)$ に分解される。次に、入力位相 $\theta_k(t)$ を求めるために、これを決定する未定係数 C_0 を求める。この未定係数 C_0 は、聴覚

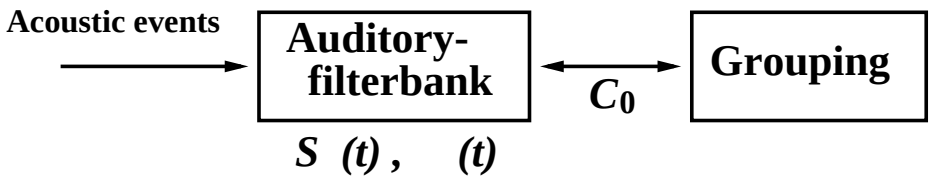

図 3.2: モデルの概要

フィルタ群とグルーピング部を仲介するようなパラメータであり、(i) 連続性の概念、(ii) 1 つの音響事象に生じる変化の物理的制約条件により一意に決定される。最後に、3 つの物理パラメータ $(S_k(t)$、$\phi_k(t)$、$\theta_k(t))$ を用いて、2 つの信号の振幅包絡 $A_k(t)$ と $B_k(t)$ を求め、すべてのチャネルについてグルーピングすることで 2 波形分離を可能にしている。

本研究では、$f_1(t)$ を純音、$f_2(t)$ を純音の周波数を中心周波数とする帯域雑音として 2 波形分離問題を議論する。

第 4 章

Gammatone filter を基底関数とした wavelet 分析合成系

この章では、聴覚フィルタのモデル化である Gammatone filter の諸特性を示すと共に、これを基底関数とした wavelet 分析合成系の設計について述べる。

4.1 Gammatone filter の特性

Patterson[29] によって設計された聴覚フィルタは、Gammatone filter と呼ばれ、基底膜の特性を模擬したものとして知られている。Gammatone filter は

$$gt(t) = At^{n-1}e^{-2\pi b_f t}\cos(2\pi f_0 t + \phi), \quad (t \geq 0) \tag{4.1}$$

のインパルス応答を持ち、ガンマ分布を表す振幅項 $At^{n-1}e^{-2\pi b_f t}$ と周波数 f_0 で表される。また、Gammatone filter の周波数特性は、$f_0 >> b_f$ のとき、

$$GT(f) \approx \left[1 + \frac{j(f - f_0)}{b_f}\right]^{-n}, \quad 0 < f < \infty \tag{4.2}$$

となる。$GT(f)$ は、$gt(t)$ 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 [29]。図 4.1に Gammatone filter の時間領域および周波数領域の特性を示す。

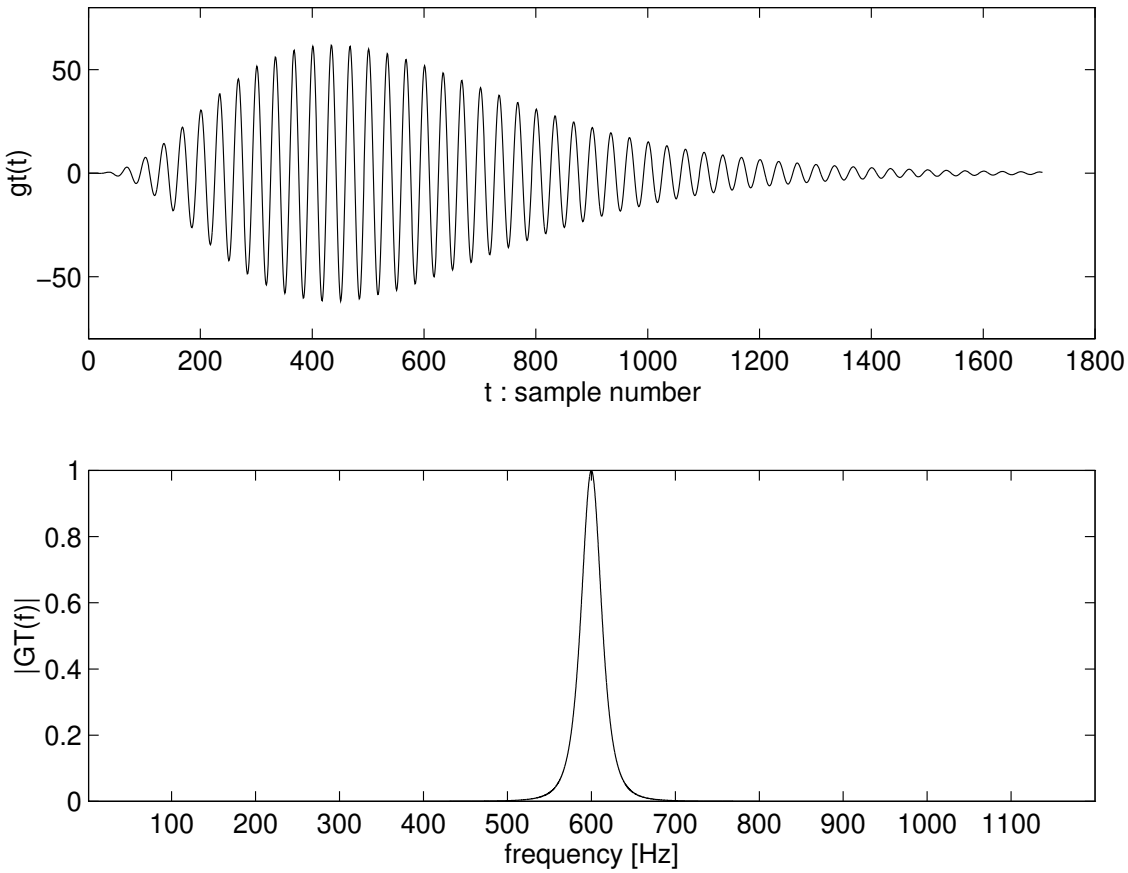

図 4.1: Gammatone filter のインパルス応答と周波数特性 (中心周波数：600[Hz]、$n = 4$,$b_f = 22.9945$)

4.2 Gammatone filter を基底関数とした wavelet 変換

従来の信号解析では、短時間 Fourier 変換 (STFT) が一般的に用いられていた。これは、中心周波数とその周波数帯域を動かしても、時間周波数の窓はその幅が変化しないという意味で固定化されたものであった。また、これは時間と周波数の不確定性原理によって束縛されたものであり、STFT は非常に高いかあるいは低い周波数を持つ信号解析に対して適切なものとはいえなかった。これに対し wavelet 変換は、時間解析窓が自動的に、高周波数現象を調べるときに狭まり、低周波数現象を調べるときに広くなるという性質を備えている。また、定 Qfilterbank による分析を数学的に記述したものとも考えられている。本研究では、

- 不連続点を検出しやすい [35]。

- 人間の内耳における信号処理と wavelet 変換に基づく信号処理に類似性がある [36]。

という利点から、wavelet 変換を利用している。

wavelet 変換で用いられる基底関数は、直交基底と擬似直交基底とに分けられるが、後者の場合、逆変換を定義するために基本 wavelet は許容条件を満たさなければならない [34]。ここで、定義 1 に連続 (積分)wavelet 変換を示す。

定義 1 (連続 (積分)wavelet 変換) 基底関数 $\psi(t) \in L^2(\mathbf{R})$ が、許容条件 *(admissibility condition)*：

$$C_\psi := \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \tag{4.3}$$

を満たすとき、ψ を基本 *wavelet* という。但し、$\hat{\psi}(\omega)$ は $\psi(t)$ の *Fourier* 変換である。基本 *wavelet* が与えられたとき、空間 $L^2(\mathbf{R})$ 上の連続 *(*積分*)wavelet* 変換を

$$\tilde{f}(a,b) = \int_{-\infty}^{\infty} f(t)\overline{\psi_{a,b}(t)} dt \tag{4.4}$$

と定義する。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、$\overline{\psi_{a,b}}$ は $\psi_{a,b}$ の複素共役である。

このとき、この逆変換 (再構成) は、定理 1 で与えられる。

定理 1 (逆 wavelet 変換 (再構成)) 関数 ψ が基本 *wavelet* で、連続 *wavelet* 変換 $\tilde{f}(a,b)$ を定義する。このとき、任意の関数 $f, g \in L^2(\mathbf{R})$ に対して

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\tilde{f}(a,b)\overline{\tilde{g}(a,b)}\right] \frac{da}{a^2} db = C_\psi < f, g > \tag{4.5}$$

が成立する。更に、関数 $f \in L^2(\mathbf{R})$ が連続である点 $x \in \mathbf{R}$ において再構成式は

$$f(t) = \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{f}(a,b)\psi_{a,b}(t)\frac{dadb}{a^2} \tag{4.6}$$

となる。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、$\overline{\psi_{a,b}}$ は $\psi_{a,b}$ の複素共役である。また、*wavelet* は、基本 *wavelet* を a 倍のスケール変換と b だけシフトしたものであり、

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \tag{4.7}$$

と表される *(*$a, b \in \mathbf{R}$*)*。

(証明)　付録 1 参照。 ■

wavelet 変換で用いられる基底関数には、Mexican-hat や Gabor 関数 [34]、Mallat による直交 wavelet[35] が知られているが、本研究では、聴覚系を模擬した聴覚フィルタ群を構築したいため、基底膜の特性を模擬した Gammatone filter を基底関数としている。

$$\psi(t) = gt(t)$$

このとき、基底関数は擬似直交基底であるが、、図 4.1に示すように $GT(0) = 0$ と式 (4.3) の許容条件を近似的に満たしているため、Gammatone filter を基本 wavelet として利用できる。便宜上、連続 wavelet 変換で議論しているが、計算機への実装には、スケールパラメータ $a = \alpha^i$、シフトパラメータ $b = j/f_s$ として離散 wavelet 変換 $(i, j \in \mathbf{Z})$ を用いている [27]。但し、f_s はサンプリング周波数である。

次に wavelet 分析合成系において位相を定義するために、実部と虚部が Hilbert 変換で結ばれるような基底関数として Gammatone filter を再定義する。

$$\Psi(t) = At^{n-1}e^{j2\pi f_0 t - 2\pi b_f t} \tag{4.8}$$

このとき複素 wavelet 変換は、定義 2 で得られる。

定義 2 (複素 wavelet 変換) 複素 *wavelet* 変換は、

$$\tilde{\mathbf{f}}(a,b) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{|a|}} f(t) \overline{\Psi\left(\frac{t-b}{a}\right)} dt \tag{4.9}$$

と定義され、振幅項 $|\tilde{\mathbf{f}}(a,b)|$ と位相項 $\arg(\tilde{\mathbf{f}}(a,b))$

$$\tilde{\mathbf{f}}(a,b) = |\tilde{\mathbf{f}}(a,b)| e^{j\arg(\tilde{\mathbf{f}}(a,b))} \tag{4.10}$$

で表すことができる *[36]*。

4.3 wavelet 分析合成系

定義 1 と定義 2 に従って、wavelet 分析合成系を作成した。本研究で作成したものは、中心周波数 $f_0 = 600$[Hz]、通過帯域を 60 ~ 6000[Hz]、フィルタ数 (チャネル数) を $N = 128$ とした。これより、スケールパラメータ $a = \alpha^i$ は、i が $-\frac{N}{2} \leq i \leq \frac{N}{2}$, α が $10^{2/N}$ となる [27]。また、シフトパラメータ $b = j/f_s$ は、サンプリング周波数 $f_s = 20$[kHz] で離散化される。ここで作成した wavelet 分析合成系の周波数特性は、各フィルタの矩形帯域幅が重複せず、図 4.2のように完全に通過帯域を被覆している [27, 28]。

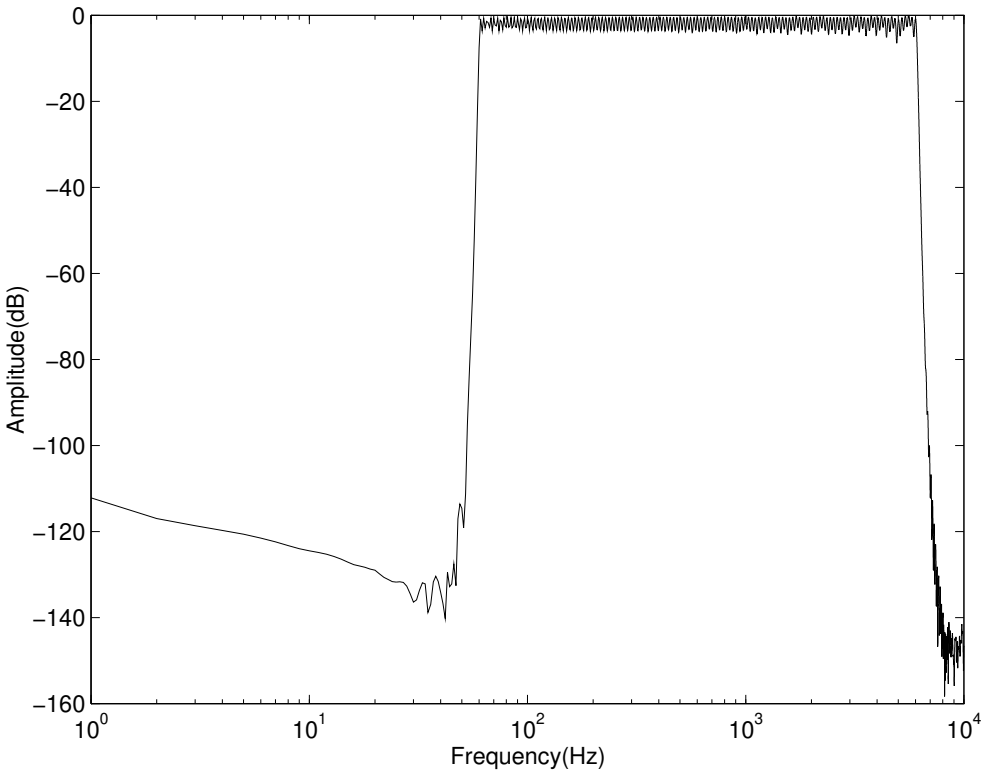

図 4.2: wavelet 分析合成系の周波数特性 ($N = 128$)

第 5 章

物理パラメータの導出方法

5.1 振幅包絡 $S_k(t)$ の計算方法

式 (3.5) における振幅包絡 $S_k(t)$ は、次の補題 1 で得られる。

補題 1 振幅包絡 $S_k(t)$ は、定義 2 の振幅スペクトル $|\tilde{\mathbf{f}}(a,b)|$ から

$$S_k(t) = |\tilde{\mathbf{f}}(a,b)| \tag{5.1}$$

で求めることができる。但し、$a = \alpha^k$、$b = t$ である。

(証明)　分析フィルタの出力 $X_k(t)$ は

$$\begin{aligned} X_k(t) &= S_k(t)\sin(\omega_k t + \phi_k(t)) \\ &= Re\{\tilde{\mathbf{f}}(a = \alpha^k, b = t)\} \end{aligned}$$

である。ここで、搬送波 $\sin(\omega_k t + \phi_k(t))$ に対し、$\cos(\omega_k t + \phi_k(t))$ の分析フィルタの出力 $Y_k(t)$ を考える。

$$\begin{aligned} Y_k(t) &= S_k(t)\cos(\omega_k t + \phi_k(t)) \\ &= Im\{\tilde{\mathbf{f}}(a = \alpha^k, b = t)\} \end{aligned}$$

これより、複素フィルタの出力 $Z_k(t)$ を

$$Z_k(t) = X_k(t) + jY_k(t)$$

とすれば、

$$\begin{aligned} Z_k(t) &= S_k(t)\sin(\omega_k t+\phi_k(t)) + jS_k(t)\cos(\omega_k t+\phi_k(t)) \\ &= Re\{\tilde{\mathbf{f}}(a,b)\} + jIm\{\tilde{\mathbf{f}}(a,b)\} \\ &= \tilde{\mathbf{f}}(a,b) \end{aligned}$$

となり、

$$\begin{aligned} |Z(t)| &= |\tilde{\mathbf{f}}(a,b)| \\ &= \sqrt{\tilde{\mathbf{f}}(a,b)\tilde{\mathbf{f}}(a,b)^*} \\ &= \sqrt{\left(S_k(t)\sin(\omega_k t+\phi_k(t))\right)^2 + \left(S_k(t)\cos(\omega_k t+\phi_k(t))\right)^2} \\ &= S_k(t)\sqrt{\sin^2(\omega_k t+\phi_k(t)) + \cos^2(\omega_k t+\phi_k(t))} \\ &= S_k(t) \end{aligned}$$

を得る。

故に、

$$S_k(t) = |\tilde{\mathbf{f}}(a,b)|$$

となる。但し、$a=\alpha^k$、$b=t$ である。 ■

5.2 出力位相 $\phi_k(t)$ の計算方法

式 (3.6) の出力位相 $\phi_k(t)$ は、次の補題 2 で得られる。

補題 2 出力位相 $\phi_k(t)$ は、定義 2 の位相スペクトル $\arg(\tilde{\mathbf{f}}(a,b))$ から

$$\phi_k(t) = \int \left(\frac{d}{db} \arg\left(\tilde{\mathbf{f}}(a,b)\right) - \omega_k \right) db \tag{5.2}$$

で求めることができる。但し、$a=\alpha^k$、$b=t$ である。

(証明) 定義 2 の位相スペクトル $\arg(\tilde{\mathbf{f}}(a,b))$ は、

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

と表され、

$$-\pi \le \arg(\tilde{\mathbf{f}}(a,b)) \le \pi$$

の区間ランプ関数となる。ここで、$-\pi$ と π に不連続点が生じているが、これは $\tan^{-1}$ の定義域による影響である。従って、不連続点は実際には連続しているものと考えてもよいため、不連続点を除く連続領域で微分すると、

$$\frac{\partial}{\partial b}\arg(\tilde{\mathbf{f}}(a,b)) = \omega_k + \frac{d\phi_k(b)}{db}$$

となる。これは、位相スペクトル $\arg(\tilde{\mathbf{f}}(a,b))$ の時間変化が、周波数 ω_k をバイアスとした出力位相の時間変化を表すことを意味している。この角周波数 ω_k は

$$\omega_k = 2\pi f_0 \alpha^k$$

であるから、

$$\frac{d\phi_k(b)}{db} = \frac{\partial}{\partial b}\arg(\tilde{\mathbf{f}}(a,b)) - \omega_k$$

となる。故に、出力位相 $\phi_k(b)$ は

$$\phi_k(b) = \int \left(\frac{d}{db}\arg\left(\tilde{\mathbf{f}}(a,b)\right) - \omega_k \right) db = \phi_k(t)$$

となる。シフトパラメータ b と時間 t は同軸上のパラメータであるため、$b = t$ である。■

5.3 入力位相 $\theta_k(t)$ の計算方法

第 2 章で説明したように、Bregman が提唱した発見的規則には、“単一音は急に変化しないし、ゆっくりと滑らかに変化する” という漸近的変化の規則がある [7]。そこで、これを定義 3 のように物理的制約条件に捉え直す。

定義 3 (漸近的変化の物理的制約条件) “振幅包絡 $A_k(t)$ の時間変化は N 回微分可能な N 次多項式で近似される”:

$$\frac{dA_k(t)}{dt} = C_N(t) \tag{5.3}$$

この制約は、$A_k(t)$ の時間変化を N 次多項式近似したものであり、これを式 (3.7) に適用することで、 1 階線形微分方程式：

$$y'(t) + \frac{P'(t)}{P(t)}y(t) = \frac{Q'(t) - C_N(t)}{P(t)} \tag{5.4}$$

が得られる。但し、$y(t)$、$P(t)$、$Q(t)$ は $y(t) = \cot\theta_k(t)$、$P(t) = S_k(t)\sin\phi_k(t)$、$Q(t) = S_k(t)\cos\phi_k(t)$ である。尚、導出過程は、付録 2 を参照されたい。

入力位相 $\theta_k(t)$ は、この微分方程式 (5.4) を解くことにより、補題 3 で得られる。

補題 3 入力位相 $\theta_k(t)$ は、微分方程式 *(5.4)* の一般解

$$\theta_k(t) = \tan^{-1}\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos\phi_k(t) + C(t)}\right) \tag{5.5}$$

として求められる。但し、$C(t)$ は各項に未定係数をもつような未定関数であり、

$$C(t) = -\int C_N(t)dt + C_0 \tag{5.6}$$

と表される。

(証明)　1 階線形微分方程式 (5.4) を解くと、

$$\begin{aligned}
y(t) &= e^{-\int\frac{P'(t)}{P(t)}dt}\left[\int\frac{Q'(t)}{P(t)}e^{\int\frac{P'(t)}{P(t)}dt}dt + C\right] \\
&= e^{-\log P(t)}\left[\int\frac{Q'(t) - C_N(t)}{P(t)}e^{\log P(t)}dt + C_0\right] \\
&= \frac{1}{P(t)}\left[\int\frac{Q'(t) - C_N(t)}{P(t)}P(t)dt + C_0\right] \\
&= \frac{1}{P(t)}\left[\int Q'(t)dt - \int C_N(t)dt + C_0\right] \\
&= \frac{Q(t) + C(t)}{P(t)}
\end{aligned}$$

を得る。但し、$C(t)$ は未定関数である。式 (5.4) より、$y(t) = \cos\theta_k(t)$ を代入することで、

$$\cot\theta_k(t) = \frac{S_k(t)\cos\phi_k(t) + C(t)}{S_k(t)\sin\phi_k(t)} \tag{5.7}$$

となり、

$$\theta_k(t) = \tan^{-1}\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos\phi_k(t) + C(t)}\right) \tag{5.8}$$

を得る。尚、この補題から、未定関数 $C(t)$ と振幅包絡 $A_k(t)$ の関係が $C(t) = -A_k(t)$ になるという結果も導き出されている。詳細については、付録 3 を参照されたい。　■

ここで未定関数 $C(t)$ を決定できれば、一意に $\theta_k(t)$ を求めることができるが、未定関数の明確な決定方法が明らかになっていない。また、最適化問題として $C(t)$ の各項の未定係数 $C_k, k = 0, 1, \cdots, N$ を推定することは可能であるが、計算量の観点から N 次の係数を推定することは、あまりにコストがかかりすぎるため不適切である。そこで、先の定義 4 を計算量の問題から定義 4 のように再定義する。

定義 4 (漸近的変化の物理的制約条件) “微小区間において、振幅包絡 $A_k(t)$ は変化しない”:

$$\frac{dA_k(t)}{dt} = 0 \tag{5.9}$$

これにより、先の 1 階線形微分方程式 (5.4) は、

$$y'(t) + \frac{P'(t)}{P(t)}y(t) = \frac{Q'(t)}{P(t)} \tag{5.10}$$

となる。制約条件の再定義により、入力位相 $\theta_k(t)$ は補題 4 で得られる。

補題 4 入力位相 $\theta_k(t)$ は、微分方程式 *(5.10)* の一般解

$$\theta_k(t) = \tan^{-1}\left(\frac{S_k(t)\sin\phi_k(t)}{S_k(t)\cos\phi_k(t) + C_0}\right) \tag{5.11}$$

として求められる。尚、C_0 は未定係数であり、未定関数 $C(t)$ の 0 次項である。

(証明)　補題 3 と同様。 ■

この未定係数 C_0 は、(i) 連続性 (近接) の概念と (ii) 1 つの音響事象に生じる変化に関する規則を物理的制約条件として考慮することで決定できる。次節では、この決定方法を説明すると共に、振幅包絡 $A_k(t)$、$B_k(t)$、入力位相 $\theta_k(t)$ との関係、および諸性質を述べる。

5.4　未定係数の決定方法

ここでは、入力位相 $\theta_k(t)$ を一意に求めるために、(i) 連続性 (近接) の概念と (ii)“1 つの音響事象に生じる変化に関する規則” を用いることで未定係数 C_0 を決定している。特に、(ii) の規則は、Bregman が提唱した発見的規則の 1 つである [7]。

はじめに、(i) の規則は、“未定係数 C_0 が何らかの境界条件で決定されるものであり、この境界においてそれぞれ分離された信号が連続性を保持しなければならない” ということを意味している。また、この発見的規則は、漸近的変化 [7] や old-plus-new の規則 [6] を加味したものである。この発見的規則を、定義 5 に示すように物理的制約条件として捉え直す。

定義 5 (連続性に関する物理的制約条件) “各物理パラメータ (振幅包絡 $A_k(t)$、$B_k(t)$、入力位相 $\theta_k(t)$) は分離境界の接合部 ($t = T_r$) で、ある幅 (ΔA、ΔB、$\Delta\theta$) で接合されていなければならない”:

$$\left|[A_k(t)]_{R:t\to+T_t} - [A_k(t)]_{R:t\to-T_r}\right| \leq \Delta A \tag{5.12}$$

$$\left|[B_k(t)]_{R:t\to+T_t} - [B_k(t)]_{R:t\to-T_r}\right| \leq \Delta B \tag{5.13}$$

$$\left|[\theta_k(t)]_{R:t\to+T_t} - [\theta_k(t)]_{R:t\to-T_r}\right| \leq \Delta\theta \tag{5.14}$$

但し、R は分離境界を表し、振幅包絡 $[\cdot]_{R:t\to+T_t}$ は、境界 R における $t = Tr$ のときの値を示す。

定義 5 に示された制約から、接合境界 R における連続性を保持した形で、未定係数 C_0 の取り得る範囲を

$$C_\alpha \leq C_0 \leq C_\beta \tag{5.15}$$

に限定できる。但し、C_α と C_β は、この境界における未定係数の上限と下限である。この概要を図 5.1に示す。基本的に、定義 5 を満たす形で、未定係数 C_0 の範囲は限定されるが、式 (5.12) ~ (5.14) すべてを満たさない場合、式 (5.12) と式 (5.13) を用いて範囲を限定できる。

次に (ii) の規則を用いる。この規則に関して、Bregman は、CMR の現象 [32] を発見的規則の 1 つである “1 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” という点で説明している [7]。この知見は、信号が同じ音源から出ている場合、分析フィルタ間の振幅包絡に強い相関があるということを意味している。そこで、この規則を帯域雑音に適用し、定義 6 に示すように物理的制約条件に捉え直す。

定義 6 (1 つの音響事象に生じる変化に関する物理的制約条件) “着目している振幅包絡 $B_k(t)$ は、図 5.2に示すような隣接する聴覚フィルタから得られる $B_{k-1}(t)$ と $B_{k+1}(t)$ に強い相関がなければならない”

$$B_k(t) \approx B_{k-1}(t) \quad B_k(t) \approx B_{k+1}(t) \tag{5.16}$$

この制約から帯域雑音の振幅間の相関が最も強くなるときの未定係数 C_0 を最適値として選ぶことで、一意に入力位相 $\theta_k(t)$ を決定できる。

$$\max_{C_\alpha \leq C_0 \leq C_\beta} Cor\left(B_k(t), \hat{B}_k(t)\right) \tag{5.17}$$

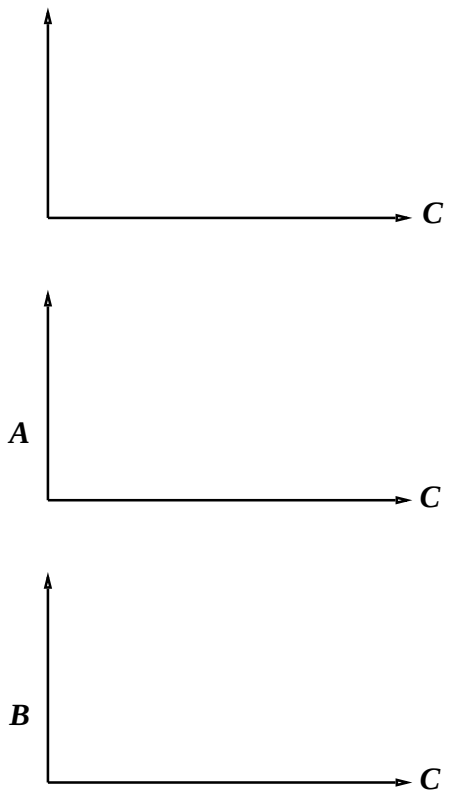

図 5.1: 未定係数 C_0 の範囲

但し、$Cor(\cdot,\cdot)$ は

$$Cor\left(B_k(t), \hat{B}_k(t)\right) = \frac{< B_k, \hat{B}_k >}{||B_k||||\hat{B}_k||} \tag{5.18}$$

であり、$\hat{B}_k = (B_{k+1} + B_{k-1})/2$ である。尚、定義 5 、 6 に関係して、未定係数 C_0 に関する諸性質が得られているが、ここでは割愛する。詳細については、付録 4 を参照されたい。

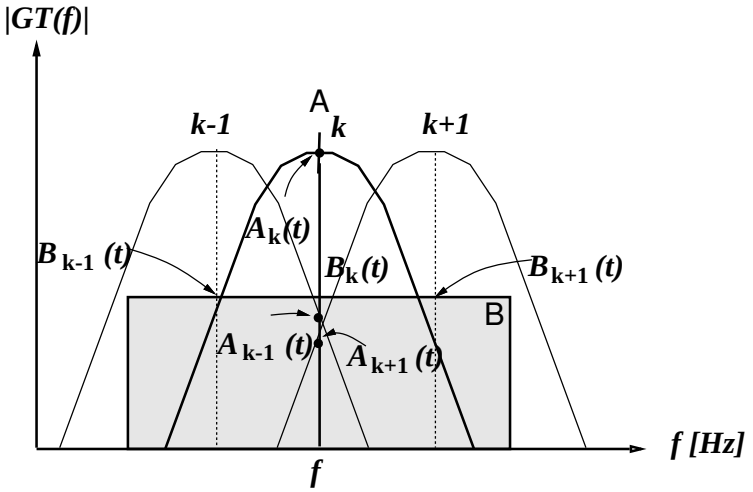

図 5.2: 隣接する聴覚フィルタの特性

5.5 2 波形分離アルゴリズム

第 5.3 節に示したように、振幅スペクトルと位相スペクトルから振幅包絡 $S_k(t)$、出力位相 $\phi_k(t)$ を求め、第 5.4 節で示した制約により入力位相 $\theta_k(t)$ を求めることで、定式化された 2 波形分離問題を解くことができる。

この節では、詳細な 2 波形分離アルゴリズムを説明する。はじめに、

1. 入力位相 $\theta_i(t), i = k-1, k, k+1$ はある区間であまり変化しない。

2. 出力位相 $\phi_i(t), i = k-1, k, k+1$ は測定可能である。

3. 振幅包絡 $A_{k-1}(t)$、$A_{k+1}(t)$ は聴覚フィルタの周波数特性から求めることができる。

という仮定を設ける。ここで、微小区間 M/f_0 において、制約条件 $\frac{dA(t)}{dt} = 0$ が成り立つものとしている。このとき、定式化された 2 波形分離問題は、次の定理 2 を用いることで解くことができる。

定理 2 (2 波形分離アルゴリズム) 定義 4、5、6 の物理的制約条件のもとで、定式化された 2 波形分離問題 *(式 (3.3) ~ 式 (3.8))* を、振幅包絡 $S_k(t)$、出力位相 $\phi_k(t)$、入力位相 $\theta_k(t)$ を用いることで解くことができる。

(証明) 補題 1、2、3、4 より明らか。 ■

この定理 2 に基づく 2 波形分離アルゴリズムの手順は、次のようになる。

1. 振幅包絡 $S_k(t)$ と位相 $\phi_k(t)$ の時間微分から検出可能な立ち上がりと立ち下がりによって全解析区間内における分離区間を決定し、
2. 更に、この分離区間を微小区間 M/f_0 に分割し、
3. 定理 2 を適用して解く。

尚、この振幅包絡 $S_k(t)$ の時間微分から、立ち上がりと立ち下がりを検出することについては、文献 [12, 13] でも支持されている。また、微小区間 M/f_0 は、隣接する聴覚フィルタ間での振幅包絡の変化を十分観測できる区間である。

この区間 M/f_0 の反復的な解析による 2 波形分離法を図 5.3に示す。この分離法では、分離区間内で区間 M/f_0 が重複しないように繰り返し解かれる。尚、図 5.3において、振幅間の相関をとるとき、境界 M/f_0 とそれより過去の境界区間すべてを利用しており、記号として境界 M/f_0 における振幅包絡を $[B_k(t)]_{M/f_0}$、それより過去を含めたものを $[B_k(t)]_F$ と表している。

1. 未定係数 C_0 の制限範囲 $C_\alpha \leq C_0 \leq C_\beta$ を決める。
2. 補題 1 と 2 から、振幅包絡 $[S_k(t)]_{M/f_0}$、出力位相 $[\phi_k(t)]_{M/f_0}$ を求める。
3. $C_\alpha \leq C_0 \leq C_\beta$ の範囲内で以下の処理をする。
   (a) ある $C_0 : C_\alpha \leq C_0 \leq C_\beta$ を仮定し、これから式 (5.11) を求めることで、2 波形の振幅包絡 $[\hat{A}_k(t)]_{M/f_0}$、$[\hat{B}_k(t)]_{M/f_0}$ を求める。
   (b) 隣接する聴覚フィルタにおいて、
      i. 隣接する聴覚フィルタの周波数特性から、振幅包絡 $[\hat{A}_{k-1}(t)]_{M/f_0}$,$[\hat{A}_{k+1}(t)]_{M/f_0}$ を求める。
      ii. 補題 1 から、振幅包絡 $[S_{k-1}(t)]_{M/f_0}$,$[S_{k+1}(t)]_{M/f_0}$ を求める。
      iii. 補題 2 から、出力位相 $[\phi_{k-1}(t)]_{M/f_0}$,$[\phi_{k+1}(t)]_{M/f_0}$ を求める。
      iv. i.–iii. から式 (3.8) を用いて、振幅包絡 $[\hat{B}_{k-1}(t)]_{M/f_0}$,$[\hat{B}_{k+1}(t)]_{M/f_0}$ を求める。
   (c) 振幅包絡 $[\hat{\hat{B}}_k(t)]_{M/f_0}$:

$$[\hat{\hat{B}}_k(t)]_{M/f_0} = \frac{[\hat{B}_{k-1}(t)]_{M/f_0} + [\hat{B}_{k+1}(t)]_{M/f_0}}{2}$$

   を求め、これより過去の情報を含む振幅包絡 $[\hat{\hat{B}}_k(t)]_F$ を求める。
   (d) $[\hat{B}_k(t)]_F$ と $[\hat{\hat{B}}_k(t)]_F$ の相関値を求める。

$$\begin{aligned} & Cor([\hat{B}_k(t)]_F, [\hat{\hat{B}}_k(t)]_F) \\ &= \frac{< [\hat{B}_k(t)]_F, [\hat{\hat{B}}_k(t)]_F >}{||[\hat{B}_k(t)]_F|| \, ||[\hat{\hat{B}}_k(t)]_F||} \end{aligned}$$

4. $Cor([\hat{B}_k(t)]_F, [\hat{\hat{B}}_k(t)]_F)$ が最大になるような未定係数 C_0 を求める。

$$\max_{C_\alpha \leq C_0 \leq C_\beta} Cor([\hat{B}_k(t)]_F, [\hat{\hat{B}}_k(t)]_F)$$

5. 補題 4 から、$[\theta_k(t)]_{M/f_0}$ を求める。
6. $[A_k(t)]_{M/f_0}$ と $[B_k(t)]_{M/f_0}$ を求める。
7. $[A_k(t)]_F$ と $[B_k(t)]_F$ を求める。

図 5.3: 2 波形分離アルゴリズム

第 6 章

2 波形分離のシミュレーション

本章では、第 5.5 節で説明した分離方法を用いて 2 波形分離が可能であることを示す。

6.1 実験データ

実験データとして、$f_1(t)$ を部分的な定常音、$f_2(t)$ を振幅変調された帯域雑音 $f_{21}(t)$ とランダム帯域雑音 $f_{22}(t)$ の 2 種類の信号を用意する。この実験データを図 6.1に示す。

$$f_1(t) = \begin{cases} A(t)\sin(2\pi f_0 t) & 0.3 \leq t \leq 0.7 \\ 0 & \text{otherwise} \end{cases} \tag{6.1}$$

$$f_{21}(t) = 1000 \sum_{f=f_0-500}^{f_0+500} B(t)\sin(2\pi f t + R(f)) \tag{6.2}$$

$$f_{22}(t) = A_C \sum_{f=f_0-500}^{f_0+500} C(f,t)\sin(2\pi f t + R(f)) \tag{6.3}$$

$$\begin{aligned} A(t) &= 1200 \\ B(t) &= 1.5 * W_m - W(t) \\ C(f,t) &= 1.5 * W_m - W(t - \Delta T \cdot R(f)) \end{aligned}$$

但し、$f_0 = 600[\mathrm{Hz}]$、$W_m = \max(W(t))$ であり、$W(t)$ は白色雑音に $30[\mathrm{Hz}]$ の低域通過フィルタをかけたものである。また、A_C は $\sqrt{f_{21}(t)^2/f_{22}(t)^2} = 1$ となる定数、ΔT は任意の定数、$R(f)$ は $[-\pi, \pi]$ の一様乱数であり、純音と帯域雑音の SN 比は $-8.51[\mathrm{dB}]$ である。

このとき、$f_2(t) = f_{21}(t)$ とした混合波形 $f_C(t) = f_1(t) + f_{21}(t)$ は CMR の場合を想定し、$f_2(t) = f_{22}(t)$ とした混合波形 $f_M(t) = f_1(t) + f_{22}(t)$ はマスキングの場合を想定している。これらの混合波形を図 6.2に示す。

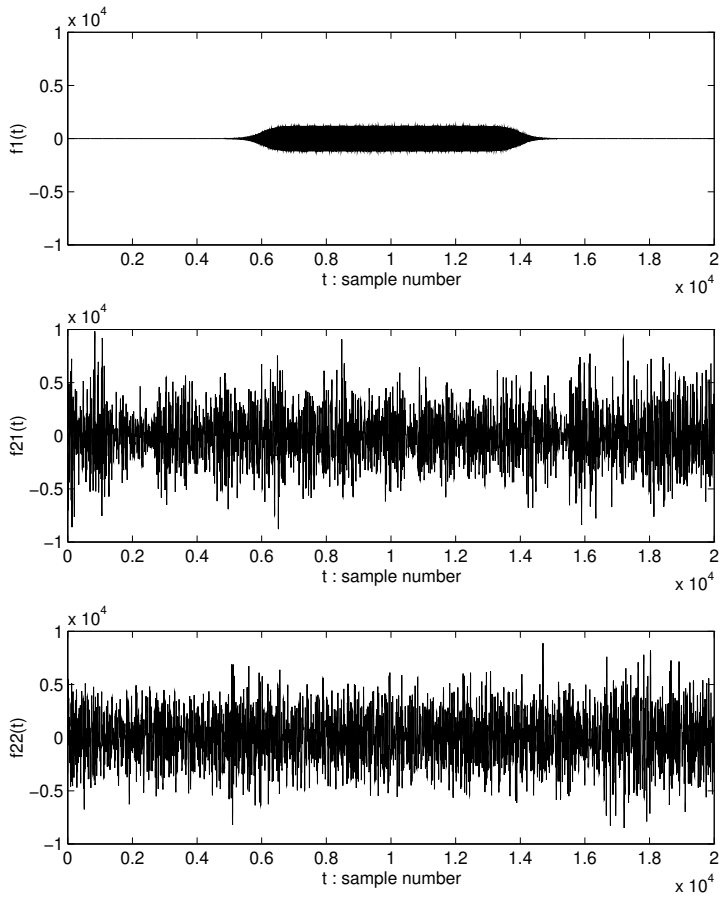

図 6.1: 音響信号：$f_1(t)$(上)、$f_{21}(t)$(中)、$f_{22}(t)$(下)

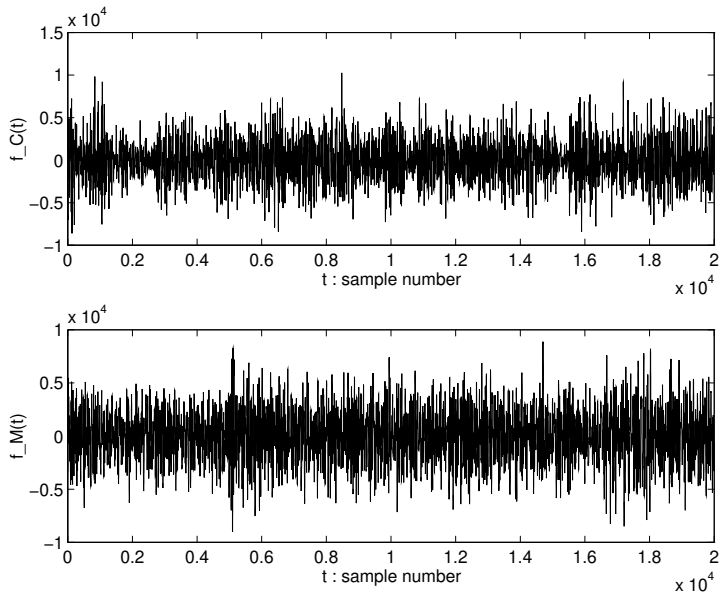

図 6.2: 混合信号：$f_C(t)$(上) と $f_M(t)$(下)

次に、ランプ関数の振幅包絡をもつ純音 $f_{11}(t)$、正弦波関数の振幅包絡をもつ純音 $f_{12}(t)$ を用意する。この実験データを図 6.3に示す。

$$f_{11}(t) = \begin{cases} D(t)\sin(2\pi f_0 t) & 0.3 \leq t \leq 0.7 \\ 0 & \text{otherwise} \end{cases} \tag{6.4}$$

$$f_{12}(t) = \begin{cases} E(t)\sin(2\pi f_0 t) & 0.3 \leq t \leq 0.7 \\ 0 & \text{otherwise} \end{cases} \tag{6.5}$$

但し、

$$\begin{aligned} D(t) &= 2000(1 + t - \frac{3}{10}) \\ E(t) &= 2000(1 + \frac{1}{10}\sin(2\pi f_c t)) \end{aligned}$$

とする。また、$f_c = 10[\text{Hz}]$ である。

このとき、$f_1(t) = f_{11}(t)$、$f_2(t) = f_{21}(t)$ とした混合信号

$$f_R(t) = f_{11}(t) + f_{21}(t)$$

は振幅変調された帯域雑音に埋もれたランプ関数の振幅包絡をもつ純音の抽出を想定し、$f_1(t) = f_{12}(t)$、$f_2(t) = f_{21}(t)$ とした混合信号

$$f_A(t) = f_{12}(t) + f_{21}(t)$$

は振幅変調された帯域雑音に埋もれた正弦波関数の振幅包絡をもつ純音の抽出を想定している。また、いずれの混合信号も共変調マスキング解除の場合を想定している。これらの混合波形を図 6.4に示す。

6.2 分離結果

はじめに、混合信号 $f_C(t)$ の分離結果を示す。混合信号 $f_C(t)$ は、wavelet 分析系により周波数分解され、分析フィルタの出力における振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ が求められる。次に、振幅包絡 $S_k(t)$ から $\frac{dS_k(t)}{dt}$ の極大点と極小点を求め、その近傍にある $\frac{d\phi_k(t)}{dt}$ の局所的な極大点 (あるいは極小点) を図 6.5のように求める。このときの局所的な極大点を立ち上がり (onset) と立ち下がり (offset) とし、分離区間 (onset – offset) を決定する。入力位相 $\theta_k(t)$ は、図 5.3の分離方法に従って未定係数 C_0 の最適値を求めることで、一意に求まる。これらの物理的手がかりを求めることで、式 (3.7)、式 (3.8) より 2 波形の振幅

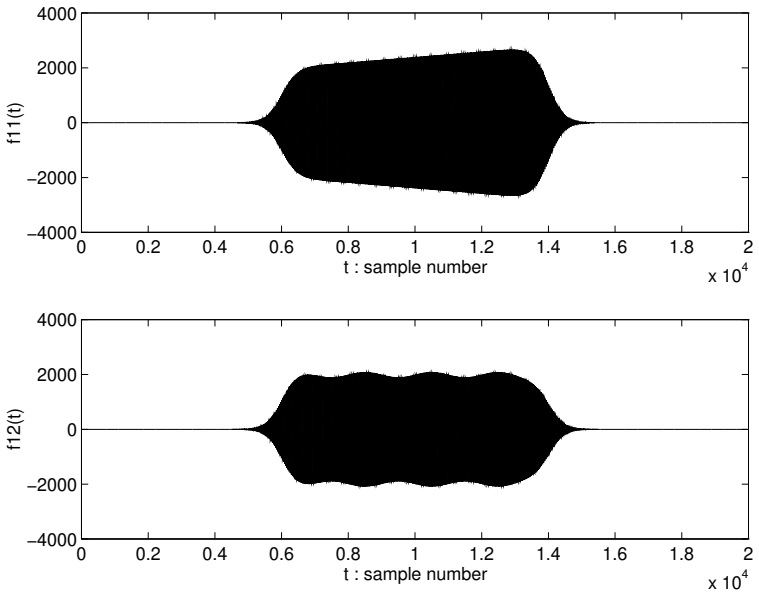

図 6.3: 混合信号：$f_{11}(t)$(上) と $f_{12}(t)$(下)

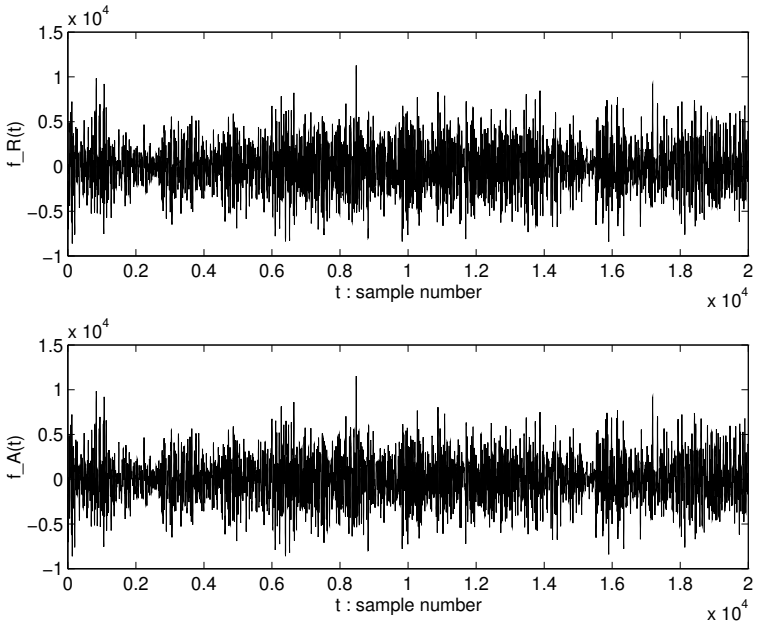

図 6.4: 混合信号：$f_R(t)$(上) と $f_A(t)$(下)

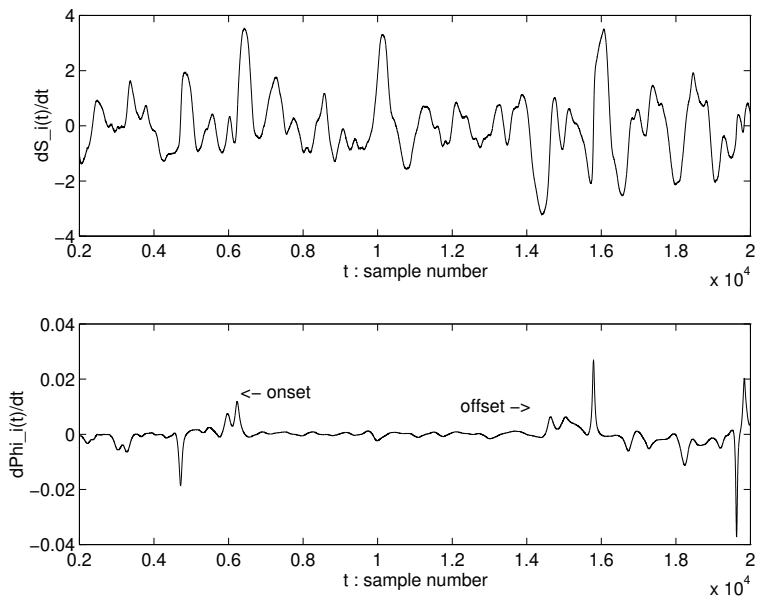

図 6.5: 振幅包絡と位相の時間微分：$\frac{dS_k(t)}{dt}$ と $\frac{d\phi_k(t)}{dt}$

包絡 $A_k(t)$ と $B_k(t)$ が計算される。中心周波数 f_0 の分析フィルタにおける振幅包絡の分離結果を図 6.7に示す。最後に、すべてのチャネルにわたって分離を行ない、それぞれの分離結果を wavelet 合成系により合成 (グルーピング) することで、$\hat{f}_1(t)$ と $\hat{f}_2(t)$ を再構成できる。ここでは、微小区間を $3/f_0$、$\Delta B = 50.0$、$\Delta\theta = \pi/20$ とし、ΔA を分離前後の $A_k(t)$ の差分値とした。この結果を図 6.8に示す。尚、$\hat{f}_1(t)$ については、周波数 f_0 が既知であるため、それを中心周波数とする分析フィルタについてだけ分離を行ない、再構成している。

同様に、混合波形 $f_M(t)$ の場合、図 6.9のように振幅包絡 $S_k(t)$ と出力位相 $\phi_k(t)$ を求め、入力位相 $\theta_k(t)$ が一意に求まることで、 2 波形の振幅包絡 $A_k(t)$ と $B_k(t)$ が図 6.10のように分離され、図 6.11のように再構成される。これらの再構成された結果をみると、振幅変調された帯域雑音の方が、ランダム帯域雑音よりも容易に純音 $f_1(t)$ を抽出できることがわかる。また、再構成された信号波形について、時間領域における SN 比：

$$SNR = 10\log_{10}\frac{\int_0^T f_i^2(t)dt}{\int_0^T (f_i(t) - \hat{f}_i(t))^2 dt}, \quad i = 1, 2, 3 \tag{6.6}$$

を用いて定量的な評価を行なったところ $f_C(t)$ のとき、$\hat{f}_1(t)$ が 15.82[dB]、$\hat{f}_2(t)$ が 11.76[dB]、$f_M(t)$ のとき、$\hat{f}_1(t)$ が 0.13[dB]、$\hat{f}_2(t)$ が 6.47[dB] であった。この結果から、混合信号 $f_C(t)$ の方が混合信号 $f_M(t)$ よりも正確に純音を抽出でき、更に分離した雑音もより正確に分離

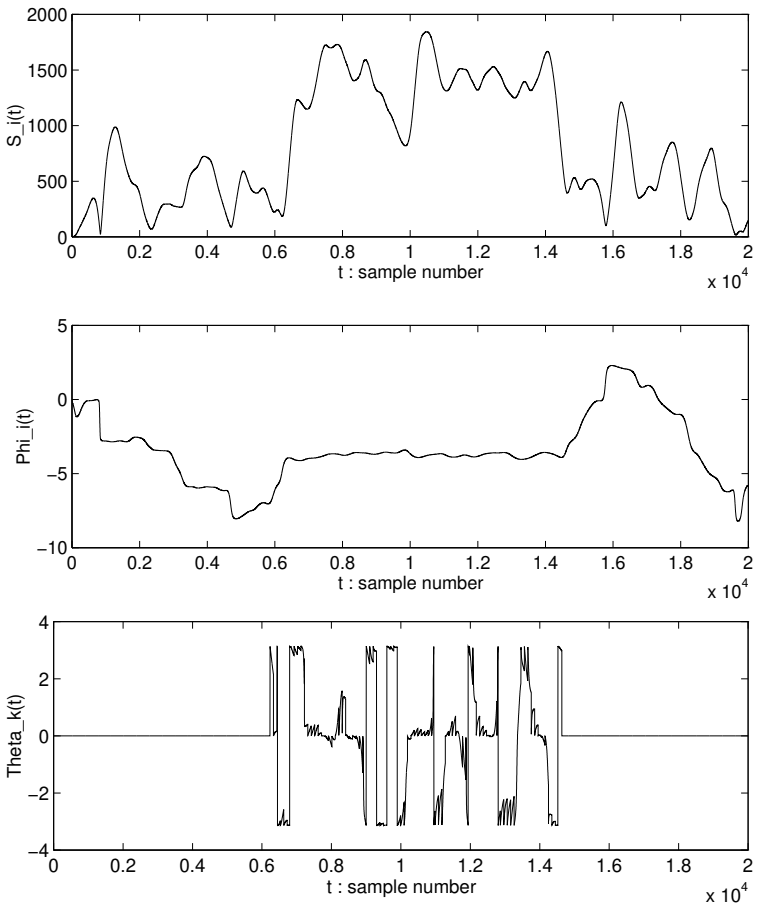

図 6.6: 混合信号 $f_C(t)$ の場合 : 振幅包絡 $S_k(t)$(上)、出力位相 $\phi_k(t)$(中)、入力位相 $\theta_k(t)$(下)

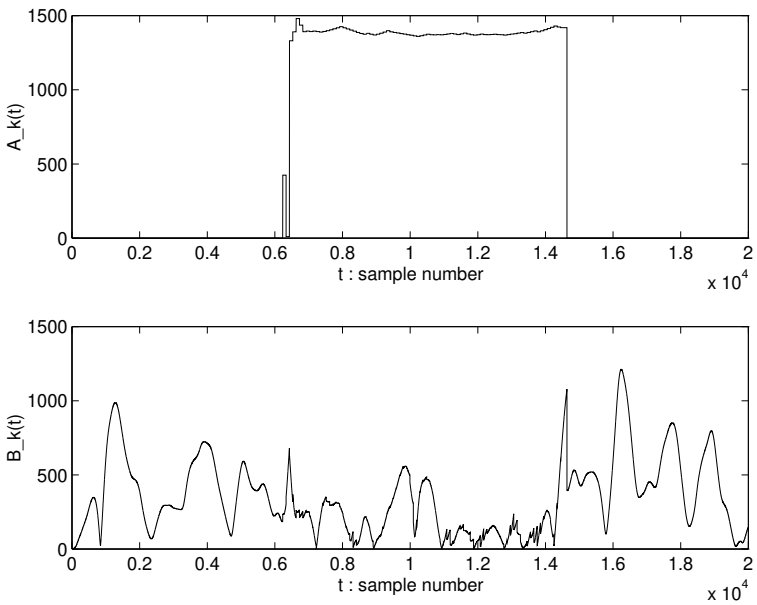

図 6.7: 混合信号 $f_C(t)$ の場合：分離された信号の振幅包絡 $A_k(t)$(上) と $B_k(t)$(下)

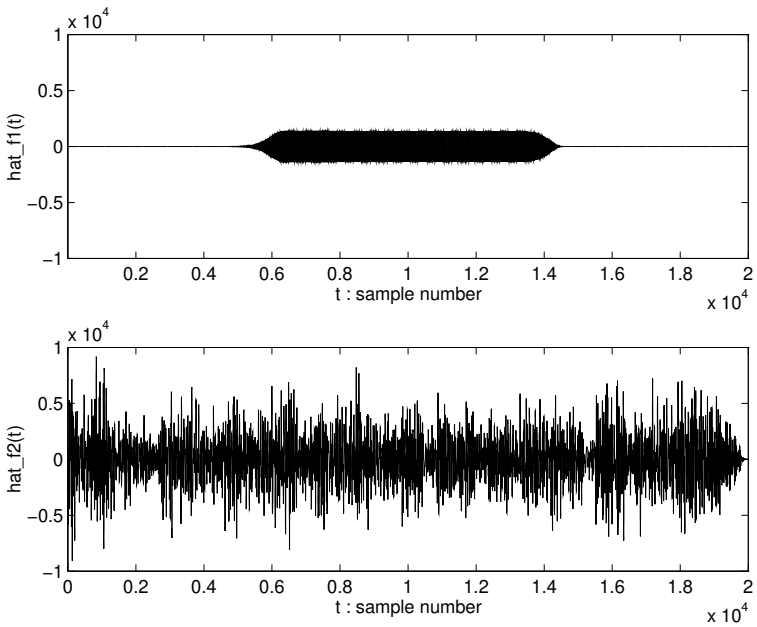

図 6.8: 混合信号 $f_C(t)$ の場合：再構成された信号 $\hat{f}_1(t)$(上) と $\hat{f}_2(t)$(下)

し、再構成できることがわかる。

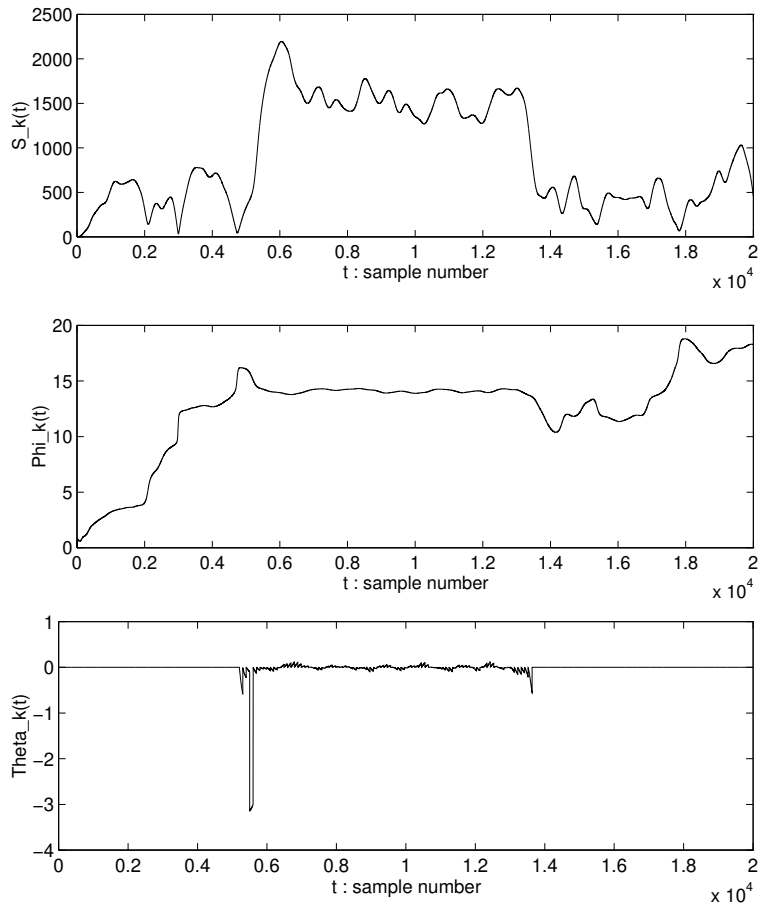

図 6.9: 混合信号 $f_M(t)$ の場合：振幅包絡 $S_k(t)$(上)、出力位相 $\phi_k(t)$(中)、入力位相 $\theta_k(t)$(下)

次に、混合信号 $f_R(t)$ と混合信号 $f_A(t)$ の分離結果を示す。混合信号 $f_R(t)$ の場合、図 6.12のように振幅包絡 $S_k(t)$ と出力位相 $\theta_k(t)$ を求め、入力位相が一意に求まることで、2 波形の振幅包絡 $A_k(t)$ と $B_k(t)$ が図 6.13のように分離され、図 6.14のように再合成される。

同様に、混合信号 $f_A(t)$ の場合、図 6.15のように振幅包絡 $S_k(t)$ と出力位相 $\theta_k(t)$ を求め、入力位相が一意に求まることで、 2 波形の振幅包絡 $A_k(t)$ と $B_k(t)$ が図 6.16のように分離され、図 6.17のように再合成される。 また、再構成された信号波形について、時間領域における SN 比を用いて定量的な評価をおこなったところ、$f_R(t)$ のとき、$\hat{f}_1(t)$ が 14.05[dB]、$\hat{f}_2(t)$ が 10.32[dB]、$f_A(t)$ のとき、$\hat{f}_1(t)$ が 14.62[dB]、$\hat{f}_2(t)$ が 10.39[dB] であった。この結果から、混合信号 $f_R(t)$ や $f_A(t)$ のようにゆっくりとした振幅包絡の変化をも

つ純音については、高い精度で抽出できることがわかる。

6.3 考察

混合波形が $f_C(t)$、$f_M(t)$ のいずれの場合も、原信号 $f_1(t)$ を雑音から浮き出した形で分離・再構成している。しかし、混合波形 $f_M(t)$ の場合、$f_1(t)$ は純音とかなり異なる信号として抽出されているのに対し、混合波形 $f_C(t)$ の場合、純音にかなり類似した信号として抽出されている。この結果は、雑音の振幅包絡 $B_k(t)$ と隣接する聴覚フィルタから得られる雑音の振幅包絡 $B_{k+1}(t)$,$B_{k-1}(t)$ に強い相関があるため、純音の抽出が容易になるものと解釈できる。

次に、混合波形 $f_R(t)$ と $f_A(t)$ の場合、前者の方が、純音を高い精度で抽出することができた。これは、純音の振幅包絡の変動が緩やかであればあるほど、分離が容易であることを示している。本モデルでは、連続性の規則と 1 つの音響事象に生じる変化に関する規則についてしか制約を入れていないが、他の規則を入れることにより、比較的速い振幅包絡の変動も抽出できるものと考えられる。

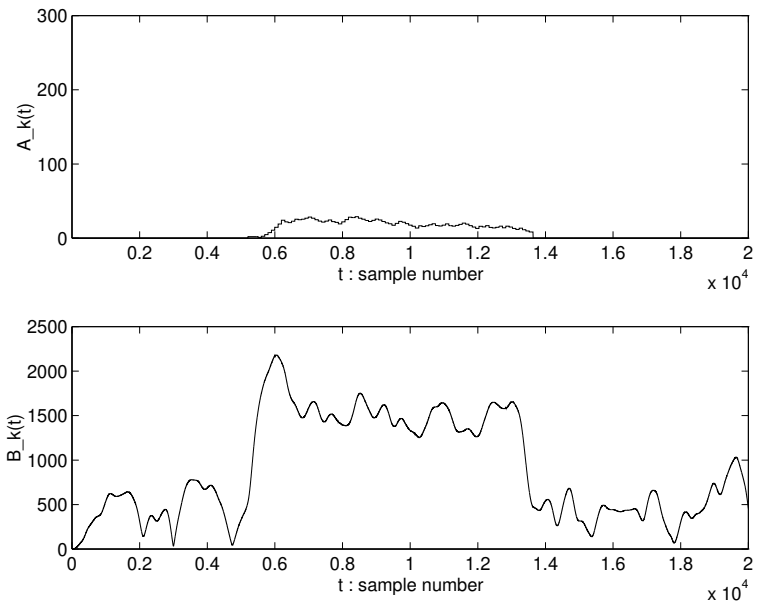

図 6.10: 混合信号 $f_M(t)$ の場合：分離された信号の振幅包絡 $A_k(t)$(上) と $B_k(t)$(下)

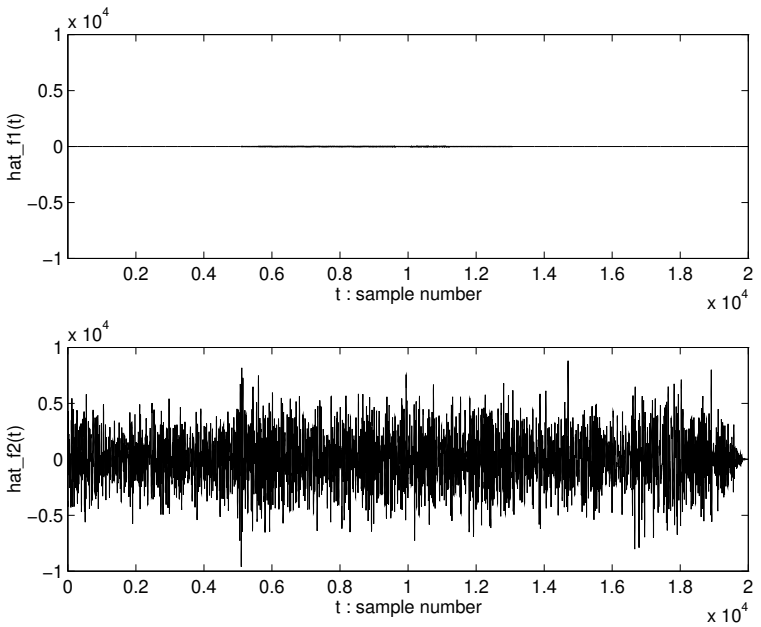

図 6.11: 混合信号 $f_M(t)$ の場合：再構成された信号 $\hat{f}_1(t)$(上) と $\hat{f}_2(t)$(下)

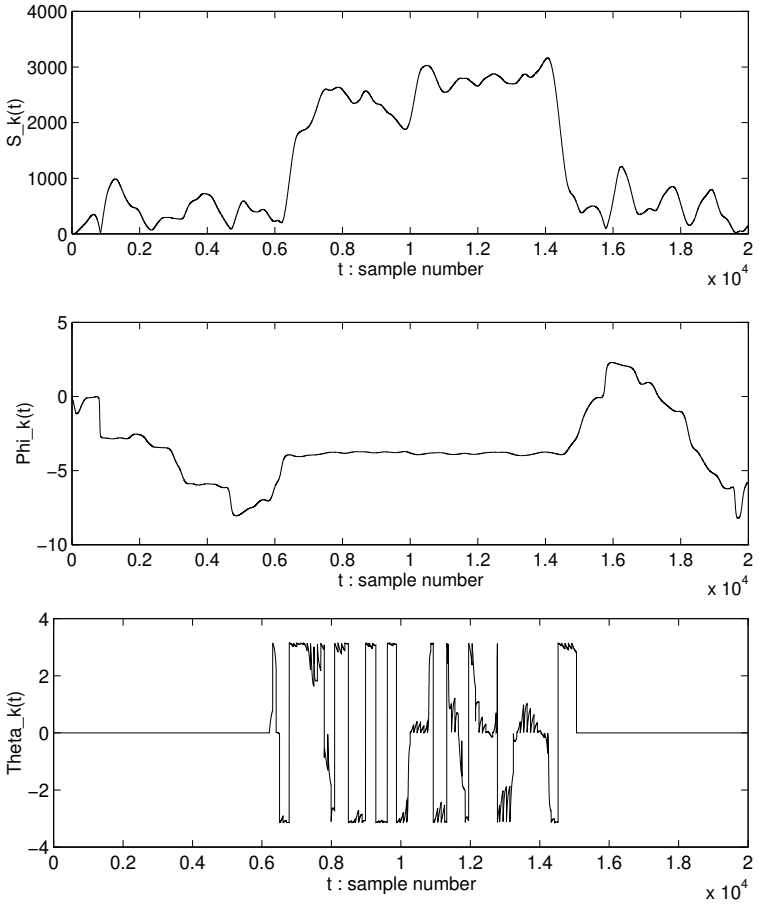

図 6.12: 混合信号 $f_R(t)$ の場合：振幅包絡 $S_k(t)$(上)、出力位相 $\phi_k(t)$(中)、入力位相 $\theta_k(t)$(下)

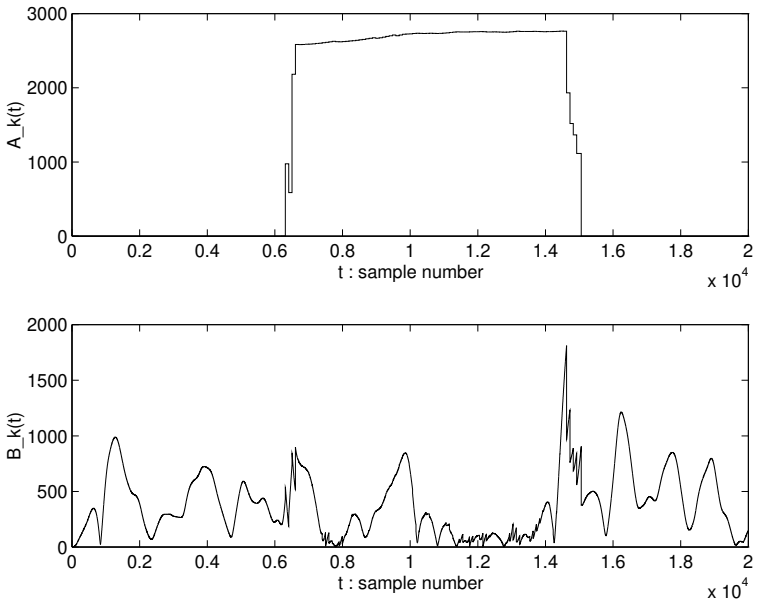

図 6.13: 混合信号 $f_R(t)$ の場合：分離された信号の振幅包絡 $A_k(t)$(上) と $B_k(t)$(下)

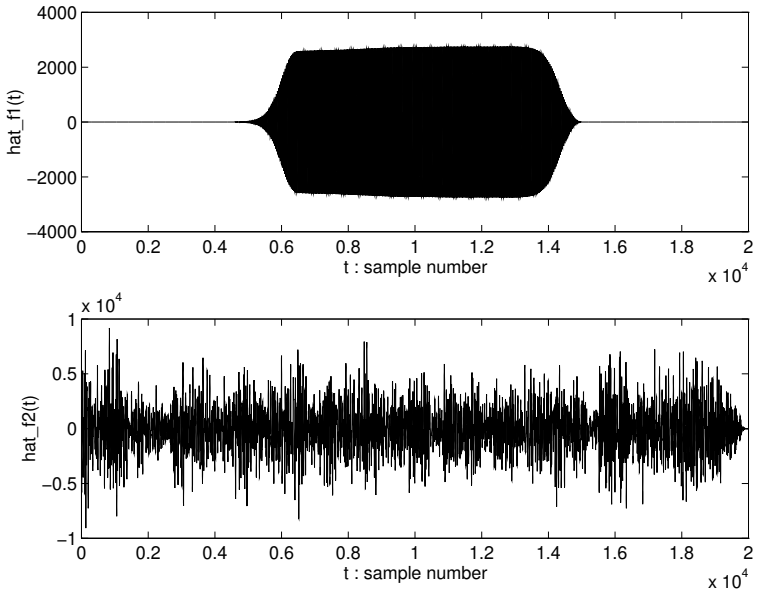

図 6.14: 混合信号 $f_R(t)$ の場合：再構成された信号 $\hat{f}_1(t)$(上) と $\hat{f}_2(t)$(下)

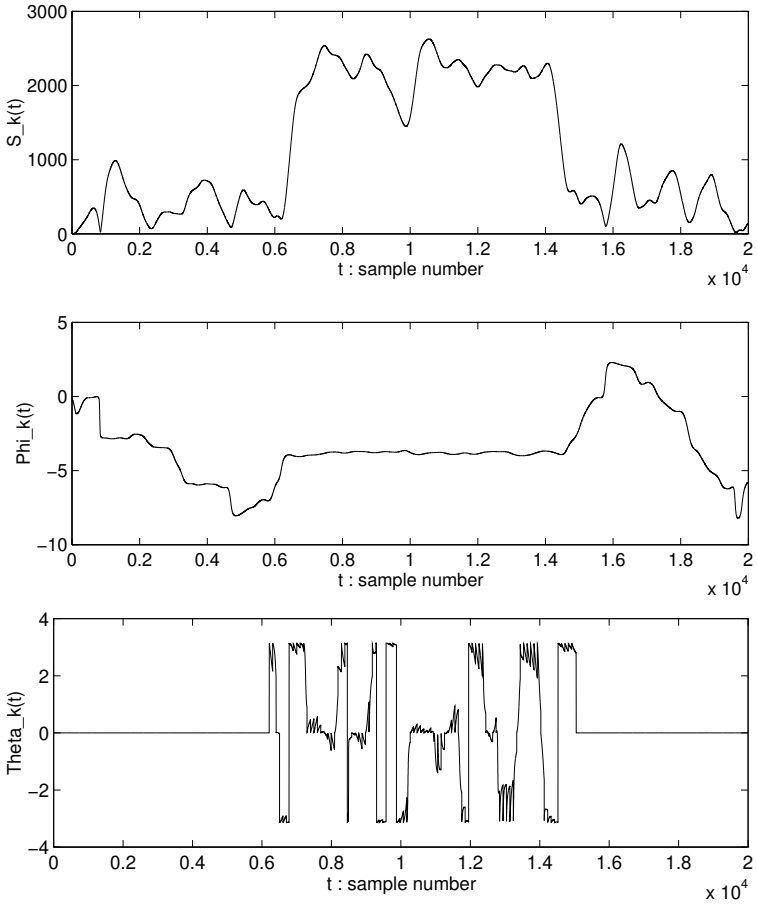

図 6.15: 混合信号 $f_A(t)$ の場合：振幅包絡 $S_k(t)$(上)、出力位相 $\phi_k(t)$(中)、入力位相 $\theta_k(t)$(下)

EPS File
FIGURE/AMSIN2.eps.epsf
not found

図 6.16: 混合信号 $f_A(t)$ の場合：分離された信号の振幅包絡 $A_k(t)$(上) と $B_k(t)$(下)

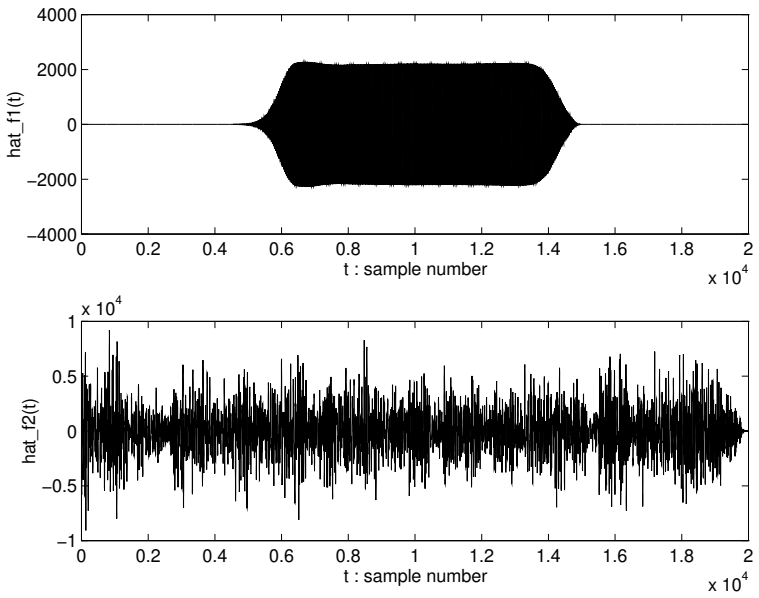

図 6.17: 混合信号 $f_A(t)$ の場合：再構成された信号 $\hat{f}_1(t)$(上) と $\hat{f}_2(t)$(下)

第 7 章

人間の聴覚系に合わせたパラメータ設定

本章では、共変調マスキング解除を説明するために、図 3.1の wavelet 分析合成系を人間の聴覚特性に合わせてパラメータ設計する。その後で、第 6.1 節で示した実験データを利用し、共変調マスキング解除のシミュレーションを行なう。

7.1 wavelet 分析合成系の再設計

聴覚フィルタの等価矩形幅 (ERB) と中心周波数の関係式は、

$$ERB(GT(f)) = 24.7\left(\frac{4.37f}{1000} + 1\right) \tag{7.1}$$

で表される [33]。この関係を図 7.1(実線) に示す。また、現段階までの wavelet 分析合成系は、聴覚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たため、聴覚フィルタの ERB は幅 1 ではなく、図 7.1の点線のような値をとっていた。図 7.1からもわかるように、周波数が 800[Hz] 付近から高域にかける領域で、特性を 1 次近似できるという報告がある [14]。本研究では、中心周波数 $f_0 = 600$[Hz] で分析合成系を設計しているので、おおよそ $f_0 = 600$ で 1 次近似できるものと仮定し、f_0 を中心に ERB(図 7.1実線) を 1 次近似することにする。

$$ERB(GT(f)) = 24.7\left(\frac{4.37f_0}{1000} + 1\right)\frac{f}{f_0} \tag{7.2}$$

この 1 次近似された ERB を図 7.1に破線で示す。この結果は、帯域幅が重複しないようにした従来の設計法で $N = 32$ としたものとおおよそ等価である。

次に、聴覚フィルタの帯域幅を 1ERB として再設計した分析合成系の周波数特性を図 7.2に示す。これからもわかるように、分析合成系の周波数は、全周波数解析 (60 ~ 6000[Hz])

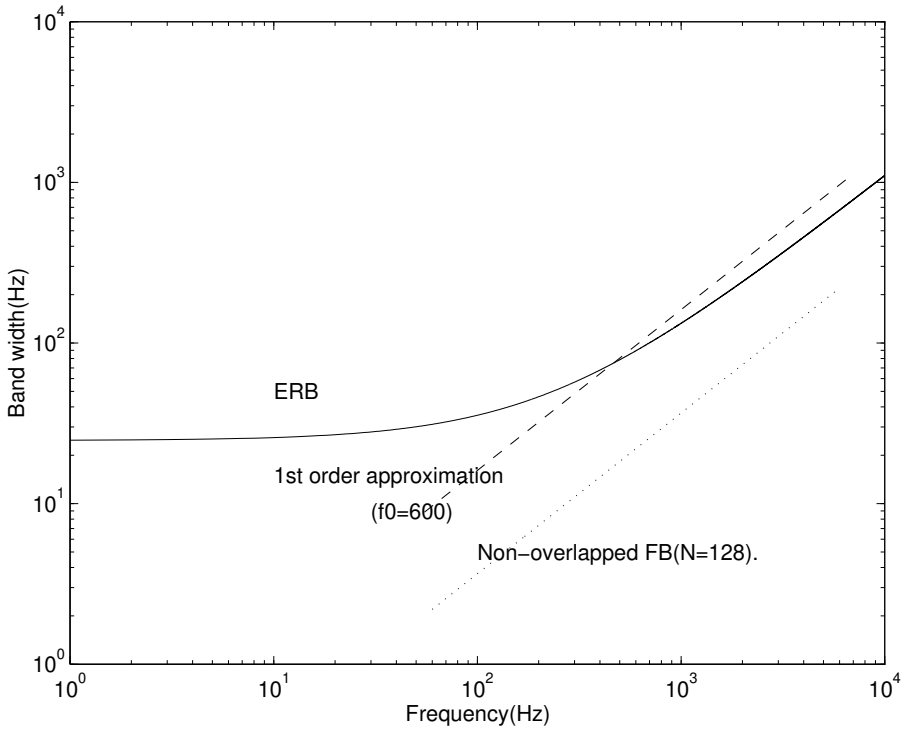

図 7.1: 周波数と ERB の関係

を完全に被覆している。また、図 7.2において、$N = 32$ のときは 1 次近似した ERB が重複しないように設計された分析合成系の周波数特性を示している。

本章では、チャネル数を $N = 128$ とし、聴覚フィルタの帯域幅が重複した状態で、かつ全解析周波数 $(60 \sim 6000[\mathrm{Hz}])$ を完全に被覆するように再設計された分析合成系を用いる。

7.2 実験データ

実験に用いるデータは、先の実験で用いた共変調マスキング解除を想定した混合信号 $f_C(t)$ とマスキングを想定した混合信号 $f_M(t)$ とした。また、隣接する聴覚フィルタの参照数を $1 \sim 8$ 個までとした。尚、純音と帯域雑音の SN 比は、$-8.51[\mathrm{dB}]$ である。

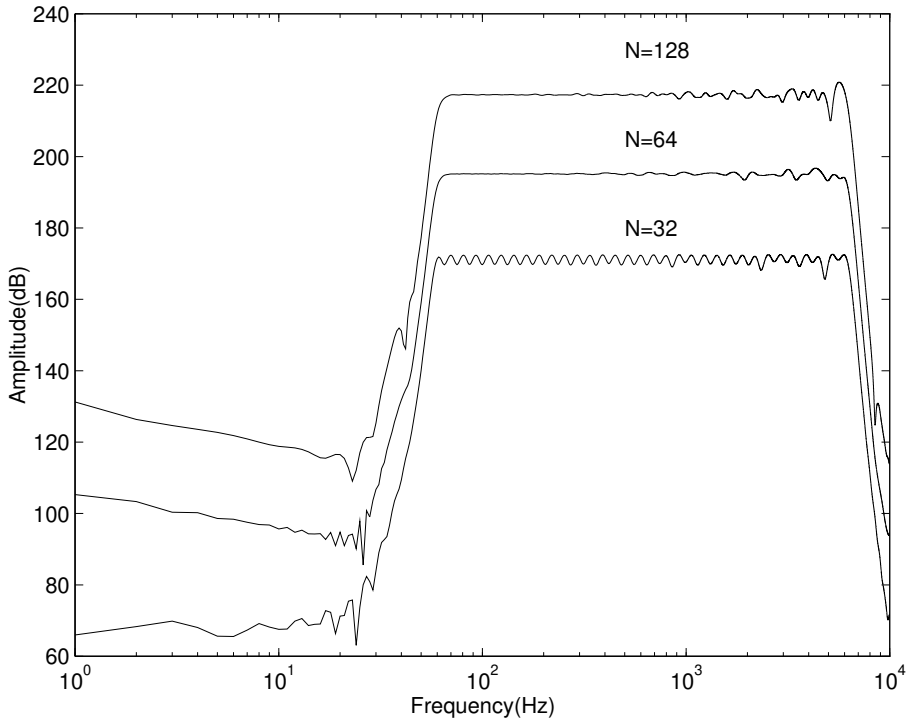

図 7.2: 再設計した wavelet 分析合成系の周波数特性

7.3 分離結果

上記の実験データに対する純音の抽出結果を表 7.1に示す。また、隣接する聴覚フィルタ数が 2 のときを基準に、相対的なマスキングの解除量を図 7.3に示す。尚、この図の縦軸は、マスキングの解除量を示すため、下向きが SN 比の向上を示している。こられの結果から、混合信号 $f_C(t)$ の場合、隣接する聴覚フィルタの参照数を増加させると、抽出された純音 $\hat{f}_1(t)$ の SN 比が向上する傾向がみられた。これに対し、混合信号 $f_M(t)$ の場合、参照数を増加させてもあまり SN 比が変わらなかった。この結果は、図 2.1と図 7.3からわかるように、振幅変調された帯域雑音の場合、帯域幅を広げることでマスキング解除が起こり、ランダム帯域雑音の場合、帯域幅を広げてもマスキング量が変らないという結果に等しいことがわかる。従って、この結果は、共変調マスキング解除の工学的な説明として解釈できる。

7.4 考察

この実験によって、共変調マスキング解除の工学的なモデル化ができた。この結果は、図 2.1に類似した結果となっているが、隣接する聴覚フィルタ数が 12 個以上のとき、純音の SN 比が若干低下していることがわかる。これは、wavelet 分析系で生じる群遅延 (高域側で狭く、低域側で広い) による影響であり、低域側における群遅延の補正を行なえば解決できる。

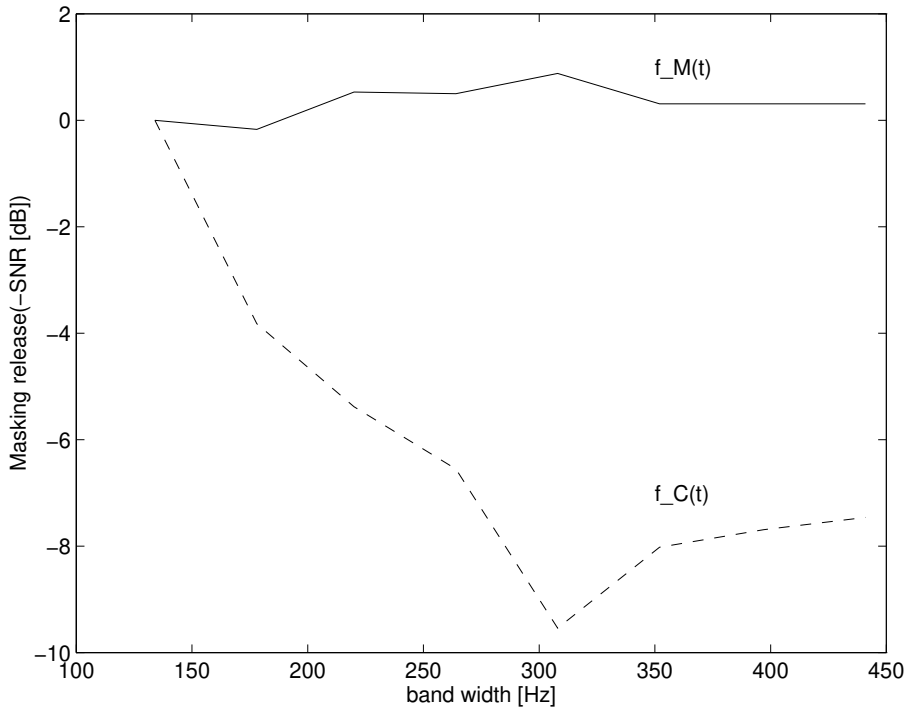

図 7.3: 帯域幅とマスキング解除の関係

表 7.1: 隣接する聴覚フィルタ数–$\hat{f}_1(t)$ の SN 比の関係

隣接する聴覚フィルタ数	帯域幅 [Hz]	$f_C(t)$ SN 比 [dB]	$f_M(t)$ SN 比 B[dB]
2	134	0.49	2.36
4	178	4.31	2.53
6	220	5.87	1.83
8	264	7.04	1.86
10	308	10.03	1.48
12	352	8.51	2.05
14	397	8.18	2.05
16	441	7.95	2.05

第 8 章

おわりに

雑音が付加された信号を wavelet 分析合成系で周波数分解し、振幅包絡と出力位相の変化及び、入力位相の変化に着目することで、原信号を抽出する方法を提案した。本方法では、振幅変調された帯域雑音が純音に付加された場合、純音の抽出が容易になり、ランダム帯域雑音が付加された場合、困難になることを示した。また、wavelet 分析合成系を人間の聴覚特性に合わせてパラメータ設定した場合、振幅変調された帯域雑音に埋もれた純音は帯域幅を広げることにより、マスキング解除が起こるのに対し、ランダム帯域雑音に埋もれた純音は帯域幅の広がりに関わらずマスキング解除が起こらないという現象が見られた。この結果は、共変調マスキング解除の工学的な説明として解釈できると共に、共変調マスキング解除の工学的なモデル化となっている。

本論文では、Bregman のいう 4 つの発見的規則 [7] のうち (i) 漸近的な変化と (ii) 1 つの音響事象に生じる変化に関する規則を物理的制約条件として取り入れたものであるが、残り 2 つ (調波関係、共通の立ち上がりと立ち下がり) についても同様に物理的制約条件として取り入れることで、雑音が付加された複合音や音声からの信号波形の抽出も可能と考えられる。

最後に、本研究で提案した抽出法において、音を分凝するためにどんな物理的手がかりを必要とし、どんな規則で解けばいいのかという必要十分条件を明らかにできれば、聴覚の計算理論の構築に向けて一歩前進できるものと思われる。

第 9 章

付録

9.1 付録 1 : 連続 wavelet 変換からの再構成

形式的に、

$$\begin{cases} F(x) & := \hat{f}(x)\hat{\psi}^*(ax) \\ G(x) & := \hat{g}(x)\hat{\psi}^*(ax) \end{cases}$$

とおく。ここで、$\hat{\psi}(\cdot)$ は $\psi(\cdot)$ の Fourier 変換であり、

$$\begin{aligned} \hat{\psi}_{a,b}(\omega) &= = \frac{1}{\sqrt{2\pi a}} \int_{-\infty}^{\infty} \psi\left(\frac{t-b}{a}\right) e^{j\omega t} dt \\ &= \frac{a}{|a|^{1/2}} \hat{\psi}(a\omega) e^{-j\omega b} \end{aligned} \tag{9.1}$$

である。また、計算の煩雑化を防ぐために、共役複素数を $\overline{\psi}$ とし、関数 $f(t)$ および $g(t)$ の wavelet 変換を $\tilde{f}(a,b)$、$\tilde{g}(a,b)$ とする。

$$\begin{aligned} & \int \left[\tilde{f}(a,b)\overline{\tilde{g}(a,b)}\right] db \\ = & \frac{1}{|a|} \int_{-\infty}^{\infty} \left\{ f(t)\overline{\psi\left(\frac{t-b}{a}\right)} dt \int_{-\infty}^{\infty} \overline{g(t)}\overline{\psi\left(\frac{t-b}{a}\right)} dt \right\} db \\ = & \frac{a^2}{|a|} \int_{-\infty}^{\infty} \left\{ \frac{1}{2\pi}) \overline{\int \overline{F}(x) e^{-jbx} dx} \right\} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{G}(y) e^{-jby} dy \right\} db \\ = & \frac{a^2}{2\pi|a|} \int_{-\infty}^{\infty} \hat{\overline{G}}(b)\overline{\hat{F}}(b) db \\ = & \frac{a^2}{2\pi|a|} \int_{-\infty}^{\infty} \overline{G}(x) F(x) dx \end{aligned}$$

となる。$F(x)$、$G(x)$ に先に値を代入し、da/a^2 について $(-\infty, \infty)$ の範囲にわたり積分すると、

$$\begin{aligned}
&\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}\left[\tilde{f}(a,b)\overline{\tilde{g}(a,b)}\right]db\right\}\frac{da}{a^2} \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\{\hat{f}(x)\overline{\hat{g}}(x)\int_{-\infty}^{\infty}\frac{|\hat{\psi}(ax)|^2}{|a|}da\right\}dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\{\hat{f}(x)\overline{\hat{g}}(x)\int_{-\infty}^{\infty}\frac{|\hat{\psi}(y)|^2}{|y|}dy\right\}dx
\end{aligned}$$

となる。ここで、

$$C_\psi = \int_{-\infty}^{\infty}\frac{|\hat{\phi}(y)|^2}{|y|}dy$$

より、

$$\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}\left[\tilde{f}(a,b)\overline{\tilde{g}(a,b)}\right]db\right\}\frac{da}{a^2} = C_\psi\frac{1}{2\pi} < \hat{f},\hat{g} >= C_\psi < f,g >$$

を得る。

以上より、

$$C_\psi < f,g >= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\tilde{f}(a,b)\overline{\tilde{g}(a,b)}\right\}\frac{dadb}{a^2}$$

の関係がいえる。最後に、$g(t)$ がデルタ関数 $\delta(t-x)$ であるとすれば、

$$\begin{aligned}
f(x) &= \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\hat{f}(a,b)\overline{< \delta(t-x),\psi_{a,b}(t) >}\right]\frac{dadb}{a^2} \\
&= \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{f}(a,b)\psi_{a,b}(x)\frac{dadb}{a^2}
\end{aligned}$$

を得る。 ■

9.2 付録 2 ：入力位相 $\theta_k(t)$ に関する 1 階線形微分方程式の導出

振幅包絡 $A_k(t)$ は、

$$A_k(t) = \frac{S_k(t)\sin(\theta_k(t)-\phi_k(t))}{\sin\theta_k(t)}$$

で表されるから、この微分は、

$$\frac{dA_k(t)}{dt} = \frac{d}{dt}\frac{S_k(t)\sin(\theta_k(t)-\phi_k(t)}{\sin\theta_k(t)}$$

$$
\begin{aligned}
&= \frac{d}{dt}\left[\frac{S_k(t)\sin\theta_k(t)\cos\phi_k(t) - \sin\phi_k(t)\cos\theta_k(t)}{\sin\theta_k(t)}\right] \\
&= \frac{d}{dt}\left[S_k(t)\cos\phi(t) - S_k(t)\sin\phi_k(t)\cot\theta_k(t)\right] \\
&= S_k'(t)\left[\cos\phi_k(t) - \sin\phi_k(t)\cot\theta_k(t)\right] + S_k(t)\frac{d}{dt}\left[\cos\phi_k(t) - S_k(t)\sin\phi_k(t)\cot\theta_k(t)\right] \\
&= S_k'(t)\cos\phi_k(t) - S_k'(t)\sin\phi(t)\cot\theta_k(t) + S_k(t)\frac{d}{dt}\cos\phi_k(t) \\
&\qquad -S_k(t)\cot\theta_k(t)\frac{d}{dt}\sin\phi_k(t) - S_k(t)\sin\phi_k(t)\frac{d}{dt}\cot\theta_k(t)
\end{aligned}
$$

となる。ここで、制約条件 $(\frac{dA_k(t)}{dt} = C_N(t))$ より

$$
S_k(t)\sin\phi_k(t)\frac{d}{dt}\cot\theta_k(t) + \left(\frac{d}{dt}S_k(t)\sin\phi_k(t)\right)\cot\theta_k(t) = \frac{d}{dt}S_k(t)\cos\phi_k(t) - C_N(t)
$$

となっている。

ここで、微分方程式を簡単に表すため、

$$
\begin{cases}
y(t) &= \cot\theta_k(t) \\
P(t) &= S_k(t)\sin\phi_k(t) \\
Q(t) &= S_k(t)\cos\phi_k(t)
\end{cases}
$$

とおけば、$\cot\theta_k(t)$ は

$$
y'(t) + \frac{P'(t)}{P(t)}y(t) = \frac{Q'(t) - C_N(t)}{P(t)}
$$

の微分方程式を解くことで求めることができる。 ■

9.3 付録 3 ：未定関数 $C(t)$ の性質

1 階線形微分方程式 5.4の一般解 $\theta_k(t)$ で用いられた未定関数 $C(t)$ は、補題 5 に示すような性質をもつ。

補題 5 未定関数 $C(t)$ は、振幅包絡 $A_k(t)$ と

$$
C(t) = -A_k(t) \tag{9.2}
$$

という関係をもつ。

(証明)　振幅包絡 $A_k(t)$ は、

$$
A_k(t) = S_k(t)\left[\cos\phi_k(t) - \sin\phi_k(t)\cot\theta_k(t)\right]
$$

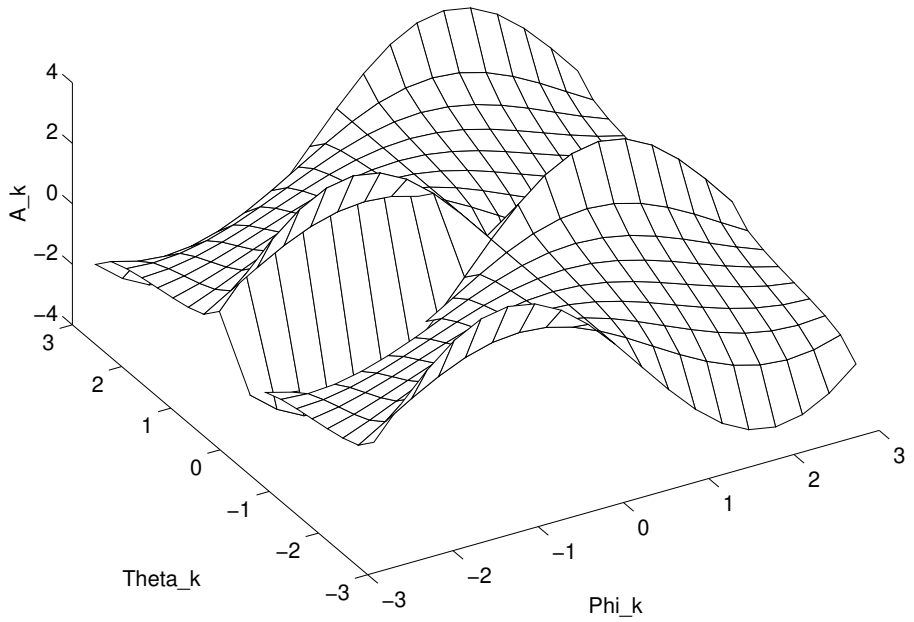

図 9.1: 各位相と振幅包絡 $A_k(t)$ の関係

で求まるため、これに入力位相 $\cot\theta_k(t)$(式 (5.5)) を代入すると

$$\begin{aligned} A_k(t) &= S_k(t)\left[\cos\phi_k(t) - \sin\phi_k(t)\frac{S_k(t)\cos\phi_k(t) + C(t)}{S_k(t)\sin\phi_k(t)}\right] \\ &= S_k(t)\left[\cos\phi_k(t) - \cos\phi_k(t) - \frac{C(t)}{S_k(t)}\right] \\ &= -C(t) \end{aligned}$$

を得る。 ■

9.4 付録４：未定係数 C_0 の諸性質

9.4.1 入力位相 $\theta_k(t)$ と出力位相 $\phi_k(t)$ に対する振幅包絡の関係

振幅包絡 $S_k(t) = 1$ とし、ある時刻 $t = T$ での入力位相 $\theta_k(T)$ と出力位相 $\phi_k(T)$ のつくる θ_k–ϕ_k 平面上で、振幅包絡 $A_k(T)$ と $B_k(T)$ がどのようなふるまいをするか調べた。この結果を図 9.1と図 9.2に示す。

この解析から次のことがいえる。

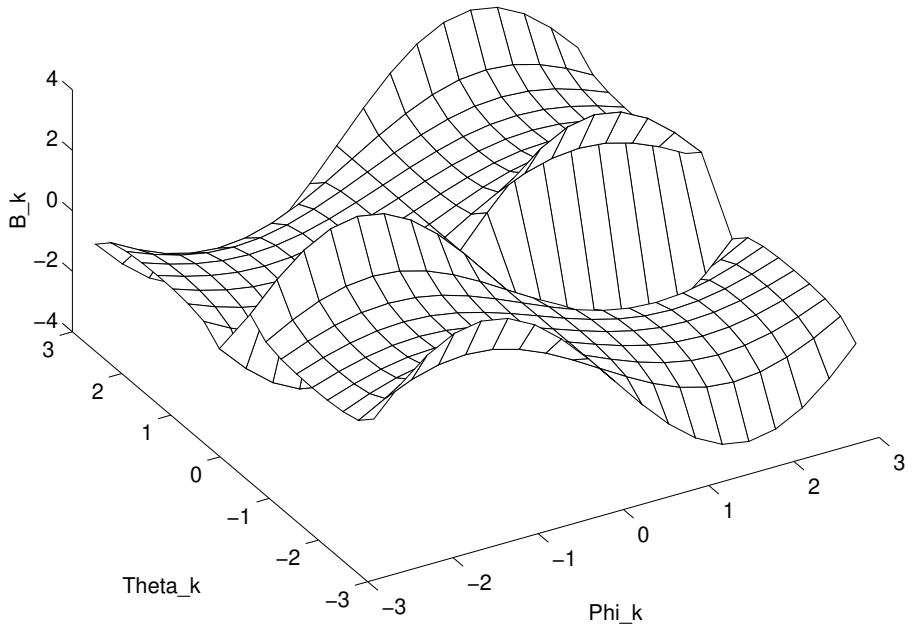

図 9.2: $\phi_k(t)$–$A_k(t)$ 特性 (上) と $\theta_k(t)$–$A_k(t)$ 特性 (下)

- 入力位相 $\theta_k(t)$ と出力位相 $\phi_k(t)$ に対する振幅包絡 $A_k(t)$ と $B_k(t)$ は、図 9.3のようなふるまい (符号の取り方) をする。
- 振幅包絡 $A_k(t)$ と $B_k(t)$ は 2 対 1 対応である。

9.4.2 出力位相 $\phi_k(t)$ と未定係数 C_0 に対する入力位相と各振幅包絡の関係

振幅包絡 $S_k(t) = 1$ とし、ある時刻 $t = T$ での出力位相 $\phi_k(T)$ と未定係数 C_0 のつくる ϕ_k–C_0 平面上で、入力位相 $\theta_k(t)$ がどのようにふるまうか調べた。この結果を図 9.4に示す。この解析結果から、次のことがいえる。

- 出力位相 $\phi_k(T) = 0$、$C_0 \leq -S_k(t)$ で入力位相 $\theta_k(t)$ に不連続点が存在する。
- 出力位相を固定した場合、未定係数 C_0 と振幅包絡 $A_k(T)$ は 1 対 1 対応である。
- 出力位相を固定した場合、未定係数 C_0 と振幅包絡 $B_k(T)$ は 1 対 2 対応である。

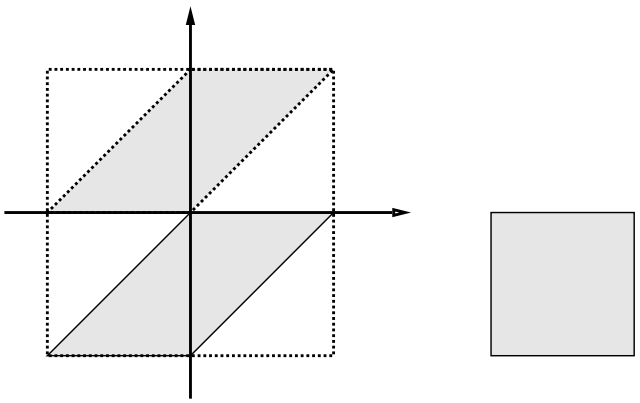

図 9.3: 各位相と振幅包絡 $A_k(t)$、$B_k(t)$ の関係

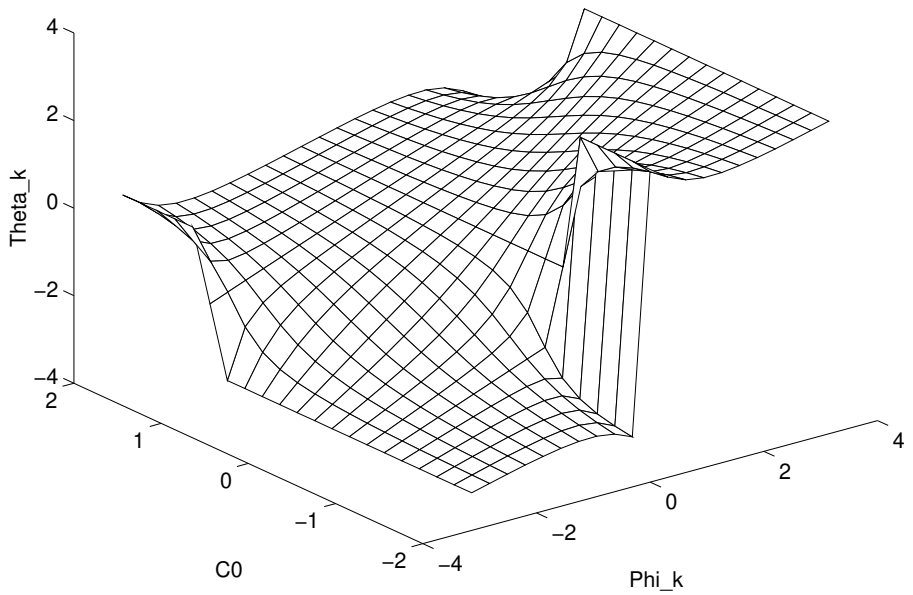

図 9.4: 出力位相 $\phi_k(t)$ と未定係数 C_0 に対する入力位相 $\theta_k(t)$ の関係

- ある未定係数 C_0 によって決定される振幅包絡 $A_k(T)$ と $B_k(T)$ は 2 対 1 対応である。

9.4.3 未定係数 C_0 と振幅包絡 $A_k(t)$、$B_k(t)$ の関係

分析フィルタの出力から得られた振幅包絡 $S_k(t)$ は

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

と表される。ここで、$A_k(t)^2$ あるいは $B_k(t)^2$ にそれぞれ $\sin^2\theta_k(t) + \cos^2\theta_k(t)$ を乗じて式変形を行なうと、

$$B_k(t) = -A_k(t)\cos\theta_k(t) \pm \sqrt{S_k(t)^2 - A_k(t)^2\sin^2\theta_k(t)} \tag{9.3}$$

あるいは、

$$A_k(t) = -B_k(t)\cos\theta_k(t) \pm \sqrt{S_k(t)^2 - B_k(t)^2\sin^2\theta_k(t)} \tag{9.4}$$

を得る。今、振幅包絡 $A_k(t) \geq 0$ を想定し、未定係数 C_0 と $B_k(t)$ の対応関係をみると、右辺第 2 項の $\sqrt{\ }$ 内にある $\sin^2\theta_k(t)$ は $-\pi/2 \leq \theta_k(t) \leq \pi/2$ の範囲内で 1 対 1 対応となるのに対し、$-\pi \leq \theta_k(t) \leq \pi$ の範囲で 2 対 1 対応となってしまう。

更に、以下のような考察もあげられる。

1. 振幅包絡 $A_k(t)$ あるいは $B_k(t)$ が実数になるためには

$$S_k(t)^2 - A_k(t)^2\sin\theta_k(t) \geq 0 \tag{9.5}$$

を満たす必要がある。

2. $-\pi/2 \leq \theta_k(t) \leq \pi/2$ の範囲では、2 波形の加算波形よりも 1 波形の振幅が大きくなることがないので、この考察から、振幅包絡 $B_k(t)$ を求めるときに、次のような範囲で解析していることがわかる。

$$\begin{cases} S_k(t)^2 \geq A_k(t)^2\sin^2\theta_k(t), & -\pi \leq \theta_k(t) < -\pi/2 \\ S_k(t) \geq A_k(t), & -\pi/2 \leq \theta_k(t) \leq \pi/2 \\ S_k(t)^2 \geq A_k(t)^2\sin^2\theta_k(t), & \pi/2 < \theta_k(t) \leq \pi \end{cases} \tag{9.6}$$

この解析範囲を図 9.5に示す。このことから、未定係数 C_0 をパラメータとして 2 波形分離問題を解くときは、入力位相 $\theta_k(t)$ の解析範囲において、

- 未定係数 C_0 に対して、$A_k(t)$ は 1 対 1 対応である、$B_k(t)$ は 1 対 2 対応である。

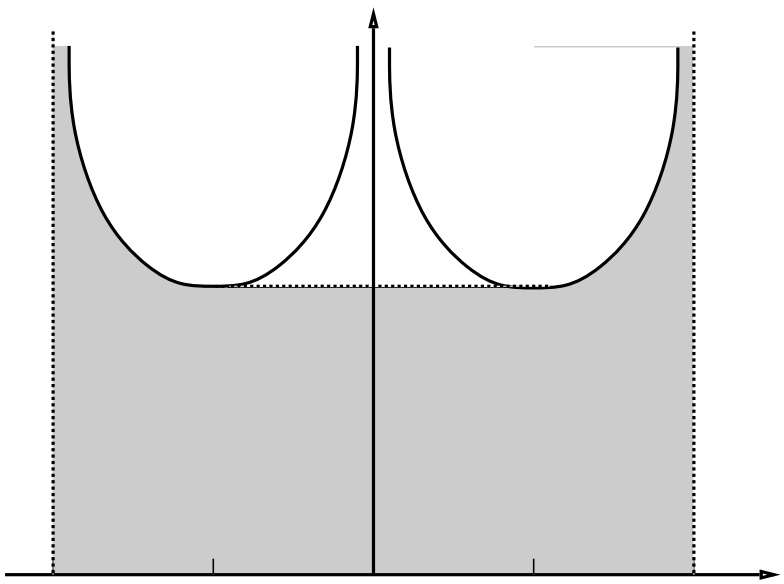

図 9.5: 振幅包絡 $A_k(t)$ の解析範囲

- 入力位相 $\theta_k(t)$ が、$-\pi \leq \theta_k(t) \leq \pi$ のとき、振幅包絡 $A_k(t)$ と $B_k(t)$ は 1 対 2 対応である。
- 未定係数 C_0 が $C_0 = -S_k(t)$ を交差するときに、入力位相 $\theta_k(t)$ は、$-\pi/2$ あるいは $\pi/2$ を交差する。

といった関係があることに注意しなければならない。

謝辞

本研究を進める上で、終始御指導して下さいました赤木正人助教授および飯島泰三教授、岩城　護助手に深甚なる感謝の意を表します。　また、研究室会議により熱心な討論並びに、有益な助言を賜わった飯島・赤木研究室の諸氏に深甚なる感謝の意を表します。

参考文献

[1] 東倉：“人間の音声処理過程と工学への応用,” 電子情報通信学会誌, vol.75, No.10 pp. 1038-1041(1992).

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

[3] 古井：“ディジタル音声処理,” 東海大学出版会 (1985).

[4] 甘利監修、中川、鹿野、東倉共著：“音声・聴覚と神経回路網モデル,” オーム社.

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

[6] A. S. Bregman : “Auditory Scene Analysis: The Perceptual Organization of Sound,” MIT Press(1990).

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

[8] 河原 : “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).

[9] 赤木：“カクテルパーティー効果とそのモデル化”, 電子情報通信学会誌 vol.78 No.5 pp.450-453(1995-05).

[10] D.Marr：“Vision,”Freeman,1992(邦訳、乾、安藤：ビジョン、産業図書 (1992)).

[11] 河原 : “聴覚の計算理論の構築に向けて,” 日本音響学会聴覚研究会資料,H-94-63(1994).

[12] 入野、Roy D. Patterson：“音響事象検出・強調の計算理論,” 日本音響学会聴覚研究会資料,H-94-64(1994).

[13] 入野：“聴覚末梢系の計算理論,” 電子情報通信学会技術報告書,SP95-40(1995).

[14] 入野：“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).

[15] M. P. Cooke : “Modelling Auditory Processing and Organization,” Ph D Thesis, University of Sheffield (1991).

[16] G. J. Brown : “Computational Auditory Scene Analysis : A Representational Approach, ” Ph D Thesis, University of Sheffield(1992).

[17] M.P.Cooke and G.J.Brown：“Computational auditory scene analysis : Exploiting principles of perceived continuity,” Speech Comunication, pp. 391-399, North Holland,13(1993).

[18] Guy J.Brown and Martin Cooke：“Comutational auditory scene analysis,” Computer Speech and Language , pp.297-336, 8(1994).

[19] D.P.W.Ellis: “A Computer Implementation of Psycoacoustic Grouping Rules,” Proc. 12th Int. Conf. on Pattern Recognition(1994).

[20] 中谷、川端、奥乃：“計算論的アプローチによる音響ストリームの分離”, 日本音響学会聴覚研究会資料、H-93-83(1993).

[21] T.Nakatani, H.G.Okuno and T.Kawabata : “Unified Architecture for Auditory Scene Analysis and Spoken Language Processing,” ICSLP’94, 24,3(1994).

[22] 中谷、後藤、川端、奥乃：“調波構造と方向同定に基づく音響ストリーム分離,” 日本音響学会講演論文集 (秋),2-3-10,(1995).

[23] 柏野、田中：“２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), J77-A, 5, pp. 731-740(1994).

[24] 柏野：“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(1994).

[25] F.Berthommier, G. Meyer: “Source separation by a functional model of Amplitude Demodulation,” ESCA. EUROSPEECH’95.4th, Madrid, September(1995).

[26] 井川、赤木：“位相の変化に着目した２波形分離法,” 日本音響学会講演論文集 (春),1994.

[27] 野田：“人間の聴覚特性を考慮した波形分離法に関する基礎的研究,” JAIST 修士論文 (1995).

[28] 鵜木、赤木：“雑音が付加された波形からの信号波形の抽出法,” 日本音響学会聴覚研究会資料,H-95-79(1995).

[29] Roy D.Patterson and John Holdsworth：“A Functional Model of Neural Activity Patterns and Auditory Images,” Advances in speech, Hearing and Language Processing, Vol3. JAI Press, London(1991).

[30] Roy D.Patterson, K.Robinson, J.Holdsworth, D.Mckown, Czhang and M.Allerhand：“Complex sounds and Auditory Images”, 9th International Symposium on Hearing:Auditory physiology and perception, June 9-14, Carcans, France(1991).

[31] Brian C.J.Moore (大串健吾監訳):“聴覚心理学概論”, 第 3 章, 誠信書房 (1994).

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

[33] 赤木：“聴覚フィルタとそのモデル”, 電子情報通信学会誌 vol.77 No.9 pp.948-956(1994-09).

[34] Chui C.K：“An Introduction to Wavelets”, Academic Press(1992).

[35] 河原、入野：“Wavelet 変換による音声の駆動源の特徴付けについて,”, 日本音響学会聴覚研究会資料,H-91-24(1991).

[36] 河原:“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).

[37] 入野：“聴覚末梢系表現からの信号再構成,”, 日本音響学会聴覚研究会資料,H-91-44(1991).

[38] “小特集「聴覚の情景解析」,” 音響学会誌,vol.50,No.12(1994).

学会発表リスト

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法,” 日本音響学会聴覚研究会, H-95-79(1995).

2. 鵜木祐史、赤木正人：“帯域雑音に埋もれた信号音の一抽出法,” 日本音響学会, 平成８年春季研究発表会,3-3-15(1996).