

Title	スペクトルピーク追跡モデルを用いたスペクトル予測 追跡に関する研究
Author(s)	坂口, 伯文
Citation	
Issue Date	1998-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/1153">http://hdl.handle.net/10119/1153</a>
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

# 修士論文

## スペクトルピーク追跡モデルを用いた スペクトル予測・追跡に関する研究

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

坂口博文

1998年2月13日

## 要旨

本論文では、聴覚の補完機能の一要因とされる周波数変化音追跡機能に着目し、音情報に対する人間の聴覚機能を Shamma らの 1 次聴覚野モデルを用いて、スペクトルピーク（周波数、振幅、帯域幅、対称性）の時間変化を予測・追跡することにより、聴覚の種々の知覚現象を再現できるモデルを提案する。

# 目次

<b>1</b>	<b>序論</b>	<b>1</b>
1.1	はじめに	1
1.2	研究の背景	2
1.3	本研究の特色	3
<b>2</b>	<b>音の知覚</b>	<b>4</b>
2.1	スイープ音知覚について	4
2.2	音声知覚について	5
<b>3</b>	<b>スペクトルピーク追跡モデル</b>	<b>8</b>
3.1	入力スペクトル	8
3.2	1次聴覚野 (A1) におけるスペクトル表現	13
3.2.1	周波数応答特性の表現	13
3.2.2	A1 における神経細胞群の応答	13
3.3	スペクトルピーク抽出アルゴリズム	14
3.4	パラメータの予測法	14
3.4.1	雑音区間の抽出	17
3.4.2	通常区間での予測機構	17
3.4.3	雑音区間での予測機構	19
3.5	複数ピークの追跡	20
3.6	4パラメータによるスペクトルの再構成	25
<b>4</b>	<b>周波数変化音予測・追跡シミュレーション</b>	<b>26</b>
4.1	複数の対数線形スイープ音に対するシミュレーション	26



4.1.1	目的	26
4.1.2	シミュレーションに使用するデータ	26
4.1.3	シミュレーション結果	29
4.1.4	考察	29
4.2	2本のクロスする対数線形スイープ音に対するシミュレーション	33
4.2.1	目的	33
4.2.2	シミュレーションに使用するデータ	33
4.2.3	シミュレーション結果	33
4.2.4	考察	33
4.3	雑音による中断のある対数線形スイープ音に対するシミュレーション	37
4.3.1	目的	37
4.3.2	シミュレーションに使用するデータ	37
4.3.3	シミュレーション結果	37
4.3.4	考察	42
4.4	合成音声に対するシミュレーション	43
4.4.1	目的	43
4.4.2	シミュレーションに使用するデータ	43
4.4.3	シミュレーション結果	43
4.4.4	考察	43
4.5	遷移部が雑音置換された合成音声に対するシミュレーション	47
4.5.1	目的	47
4.5.2	シミュレーションに使用するデータ	47
4.5.3	シミュレーション結果	48
4.5.4	考察	48
<b>5</b>	<b>結論</b>	<b>51</b>
5.1	まとめ	51
5.2	今後の課題	52
	謝辞	53
	学会発表リスト	56

# 目 次

2.1	調音結合 (実線: 連続発話のホルマント、点線: 単独発話のホルマント)	6
3.1	スペクトルピーク予測追跡モデルのフローチャート	9
3.2	両方法から得られるスペクトル (実線: 不偏推定法 (unbias)、破線: 改良ケプストラム法 (impcep)、点線: FFT スペクトル)	10
3.3	バイアスの比較 (実線: 不偏推定法 (unbias)、点線: 改良ケプストラム法 (impcep))	10
3.4	入力スペクトル (男性: /あ/)	12
3.5	自己回帰直線を用いたピーク抽出 (実線: スペクトル包絡、鎖線: 自己回帰直線)	15
3.6	A1 表現からのピーク抽出	16
3.7	実効値を用いた雑音の判断 (○印: 実効値、実線: 閾値)	18
3.8	パラメータの予測結果: シミュレーションデータの概図 [上段]、ピークの周波数 [中段]、対称性の予測 [下段] (実線: 予測追跡結果、○印: 抽出されたピーク情報)	21
3.9	パラメータの予測結果: シミュレーションデータの概図 [上段]、ピークの振幅 [中段]、帯域幅の予測 [下段] (実線: 予測追跡結果、○印: 抽出されたピーク情報)	22
3.10	複数ピークの追跡	23
3.11	本モデルを用いた複数ピークの予測・追跡の例 (実線: 予測追跡結果、破線: 原信号のピーク周波数)	24
3.12	スペクトル再構成結果 (実線: 再構成スペクトル、破線: 原信号スペクトル)	25
4.1	シミュレーションデータ 1	28

4.2	シミュレーションデータ 2	28
4.3	ピーク周波数の予測・追跡結果 ( 実線 : 予測追跡結果、破線 : 原信号のピーク周波数 )	29
4.4	入力スペクトル [上]、予測追跡されたスペクトル [下]	30
4.5	リングングの制御 ( 実線 : $f_n=5$ Hz, $df=1$ 、破線 : $f_n=8$ Hz, $df=0.63$ 、点線 : 原信号のピーク周波数 )	31
4.6	入力スペクトル [上]、予測追跡されたスペクトル [下]	32
4.7	クロスするスイープ音 ( Xパターン )	34
4.8	バウンド効果 ( Vパターン )	34
4.9	交差するスイープ音の入力スペクトル	35
4.10	交差するスイープ音の出力スペクトル	36
4.11	雑音置換されたスイープ音 1	38
4.12	雑音置換されたスイープ音 2	38
4.13	雑音置換されたスイープ音の入力スペクトル	39
4.14	雑音置換されたスイープ音の出力スペクトル	40
4.15	雑音置換内で変化方向の変わる信号のピーク周波数の予測追跡 ( 実線 : 予測追跡結果、点線 : 抽出されたピーク情報 )	41
4.16	連続合成母音 /u/-/a/-/u/	44
4.17	ピーク周波数の予測追跡結果 ( 実線 : 予測追跡結果、破線 : 原信号のピーク周波数、点線 : 音韻性の良い/a/のピーク周波数 )	44
4.18	連続合成母音の入力スペクトル	45
4.19	連続合成母音の出力スペクトル	46
4.20	遷移部が雑音置換された合成母音 /a/-/i/	47
4.21	ピーク周波数の予測追跡結果 ( 実線 : 予測追跡結果、点線 : 原信号のピーク周波数 )	48
4.22	遷移部が雑音置換された合成母音 ( 入力スペクトル )	49
4.23	遷移部が雑音置換された合成母音 ( 出力スペクトル )	50

# 表 目 次

4.1 シミュレーション条件 . . . . .	27
--------------------------	----

# 第 1 章

## 序論

### 1.1 はじめに

人間は、カクテルパーティ効果と呼ばれる複数の音源から到来する音の中から特定の音源の音だけを抽出したり、音韻修復と呼ばれる物理的には存在しない音を補って聞く機能を持っている。これらの機能のおかげで、音情報が雑音などによってマスクされたり、欠落するような場合でも、その情報を補完し正確なコミュニケーションができる。

そこで、この補完機能の一要因とされる周波数変化音追跡機能に着目し、音情報に対する人間の聴覚機能を信号処理レベルにおいて実現することによって、雑音下や複数の音源の存在する状況下における音声認識や音声合成などの様々な音声情報処理へ応用することができる。

以下第 1 章では、本研究の背景と特色を、第 2 章では音の知覚についてバウンド効果、音韻修復現象などの例を取り上げて説明する。第 3 章では周波数変化音追跡モデルについての概要、および部分的な詳細を説明する。第 4 章では複数ピークからなる様々な周波数変化音、およびホルマントシンセサイザによって作成した合成音、またそれらの音が雑音置換された場合などを対象にしたシミュレーション結果を示す。第 5 章では、本研究のまとめと今後の課題を考察する。

## 1.2 研究の背景

1990年にBregman[1]によって、心理物理学的知見から、聴覚の初期段階における音の音脈化(streaming)や群化(grouping)に関する発見的規則が発表された。これを機に、一連の音脈を1つのまとまったものとして捉えるグルーピングの効果とその形成機構に関する研究が盛んになった。

Cookeら[2][3]はオンセット、オフセットの同時性、調波性及び周波数変化の連続性に基づいた分離を、中谷ら[4]は調波性と音源方向推定を用いた分離を、Ellisら[5]は調波性、オンセットの同時性及び周波数変化の連続性に基づいた分離を試みている。

また、松井ら[6]は調波成分が同質の2つの周波数変化音がクロスする場合、周波数が交わらないバウンドする知覚が得られるバウンド効果や同方向へ変化する際、新たに音脈を形成することを心理物理実験で確認している。

倉片ら[7]は、周波数変化音に対する連続聴効果に関する検討を行なっている。連続聴効果とは、例えば強さの異なる2つの音が交互に呈示された場合、それらの2種類の音はそれぞれ別の音脈として知覚されるような現象である。倉片らは途中に雑音を置換した様々な変化パターンのスイープ音を用いて、その軌跡の知覚について心理物理実験を行なった。その結果より、雑音呈示中に知覚される音の軌跡は、雑音の開始点と終了点の周波数を単純に結んで補完するのではなく、さらに複雑な連続知覚が行なわれており、その一要因として周波数変化を予測・追跡する聴覚機能が存在する可能性を報告している。

周波数変化音の追跡機能に関する研究として、相川ら[8][9]は対数線形スイープ音に対する知覚を対象とした実験から、スイープ音の追跡モデルは2次系の伝達特性を考慮することで実現可能であることや、神経演算モデルを用いて複数の周波数変化音を追跡し、種々の周波数変化音の知覚現象をモデル化している。

一方、増田ら[10]は連続母音の過渡部の終端を雑音置換することによって、過渡部が延長したと知覚されることを心理物理実験により調べ、ホルマント追跡系の存在を示唆している。しかし、相川らのモデルは周波数のみの追跡であり、増田らの手法には時間的接続に問題がある。

そこで、河本[11]は、聴覚末梢系や聴神経の信号処理機能に関する生理学的知見を考慮した、Shammaら[12]による1次聴覚野(Auditory Cortex 1:A1)における知覚モデルに基づき、周波数変化音追跡機能モデルを作成した。しかし、このモデルは単一ピークしか追跡できないこと、ピーク抽出精度が低い、内挿機能は含まれていないなどの問題が

ある。

そこで、複数ピーク抽出・追跡アルゴリズムを考案し、複数のスペクトルを追跡する機構をこのモデルに加え、音韻修復現象などの補完現象を工学的に表現するモデルを構築することが本研究の目的である。

### 1.3 本研究の特色

本研究では、スペクトルピークを振幅、周波数、帯域幅、対称性の4パラメータで表現する。そこで、Shammaらによって提案された、聴覚末梢系の信号処理機能に関する生理学的知見を考慮した1次聴覚野(A1)におけるスペクトル知覚モデルを用いている。

スペクトルピークを表現する4つのパラメータ(周波数、振幅、帯域幅、対称性)の時間変化を2次系システムを用いて予測し、Bregmanの発見的規則の一つで、音脈の形成に関わる規則である、「よい連続の原理や周波数近傍の変理の概念」に従って、任意の時刻における複数ピークの追跡を行なう。

本モデルを用いることによって、人間の聴覚の知覚現象であるバウンド効果を模擬し、雑音の重畳した信号から目的の信号を抽出することにより音韻修復現象を模擬できる。

## 第 2 章

# 音の知覚

人間の聴覚における音の知覚について、過去の研究を考察しながら説明する。

### 2.1 スイープ音知覚について

周波数が時間と共に変化する周波数変化音 (FM 音) の知覚に関する研究は古くから行なわれており、周波数変化の検知限については詳細に調べられてきた。

相川らは対数線形スイープ音を刺激音として用いた心理物理学実験から、

- 周波数が対数的に上昇、下降する刺激音が湾曲して知覚される。
- 急激に周波数変化率が変わる一方向性周波数変化音によりピッチのリングング (振動) が知覚される。リングングは追跡系の時間差によって引き起こされるオーバーシュートによるものである。

ことを調べ、この 2 つの現象が、低い固有周波数と高い固有周波数を持つ 2 次系の伝達関数で周波数追跡過程を近似することによって説明できることを報告している。

一方、低次の聴覚系は基本的にはバンドパスフィルタとして機能し、周波数変化音は複数のバンドパスフィルタの出力に分解されることになる。それを一連の音脈として統合するためには、更に上位の処理が必要とされる。この処理過程の動的特性は周波数変化音知覚と密接な関係があると考えられている。

Bregman により提唱された聴覚情景解析により、一連の音刺激を一つのまとまりとしてとらえる音脈化 (streaming) あるいはグルーピングの効果と、その形成機構に関する研



究が行なわれるようになった。その中で、調波成分が同質の2つの周波数変化音がクロスする場合、周波数が交わらない「バウンド」する知覚が得られると報告している。この知覚現象も周波数変化音を予測追跡することに起因していると考えられる。

松井らは心理物理実験によって、2つの周波数変化音がクロスする場合バウンド効果が知覚されることを確認し、同時的音脈分析層、継時的音脈分析層などから構成される心理的過程の予測処理モデルを提案している。また、下降変化する2つの周波数変化音はバウンドするのではなく、知覚像が融合され新たな音脈を形成すると報告している。

また倉片らは、聴覚情景解析の観点から、周波数変化音に対する連続聴効果についての心理物理実験を行なった結果から、刺激音に雑音を後続させた場合

- 雑音を後続させなかった場合に比べ刺激音の持続時間が過大評価される。更に、後続雑音の持続時間が長くなるに従い過大評価量が多くなる。
- 下降(上昇)変化する刺激音の終端周波数が、雑音を後続させなかった場合に比べ、より低く(高く)知覚される。

これらのことを調べることにより、聴覚には雑音前の音の情報を用いて周波数変化音を予測追跡することによって、雑音中を外挿する機能が存在することを示した。またその機能は、雑音の開始点と終了点の周波数を結んで単純に補完するのではなく、複雑な連続性の知覚が行なわれていると報告している。

## 2.2 音声知覚について

人間は音声を知覚する場合、その有力な情報の一つにスペクトルピークの時間的な遷移、すなわち周波数の時間的な変化情報が挙げられる。音素を特徴づける優勢な周波数成分は声道の共振周波数(formant)に対応しており、低い方から第1、第2、第3ホルマントと呼ばれる。特に、母音は3個程度の特徴的なホルマントがあり、単独に発生された母音に対しては第1、第2ホルマントの2つでほぼその音響的な特徴を決定する事ができる。

一方、連続音声の中では声道の形が急に变化できないことから、前後の音素の影響を受け、ホルマント周波数の値が単独に発生された時の値とずれる性質がある。このような音素と音素の中間部において、音響的性質が連続に推移する現象は調音結合と呼ばれる。調音結合によって、音響的特徴が不完全な音が生じるため、ホルマントだけでは音の区別を

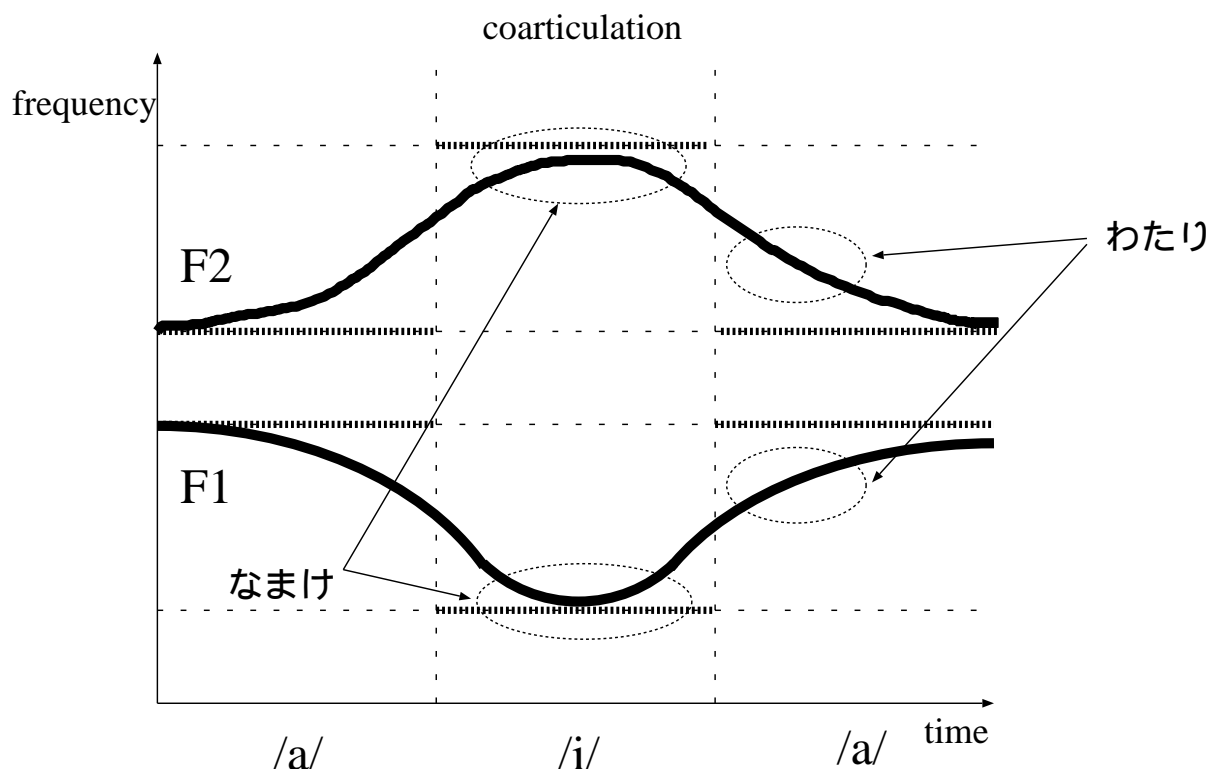


図 2.1: 調音結合 ( 実線 : 連続発話のホルマント、点線 : 単独発話のホルマント )

することが難しくなる。音響的な不完全音には、音声の物理的特徴が単独発話されたときの物理的特徴に到達する前に、次の音に移る「なまけ音」や、音韻から音韻への遷移状態に他の音韻の物理的特徴を通過する「わたり音」がある。( 図 2.1 )

人間がこれらの不完全な音響的特徴を知覚する際には、聴覚の予測追跡機能のような補正機能が存在し、動的音響特徴から不完全な音響的特徴を補正して知覚されると考えられる。

また、スweep音と同様に、時間的に遷移する音声の一部が雑音などによってマスクされたり、欠落している場合など、音韻修復と呼ばれる物理的には存在しない音を補って知覚する。この現象の一要因は、聴覚の時間的に変化する周波数を予測追跡する機能によるものと考えられる。

このように、人間は音韻を特徴付ける要因の一つに時間的に変化する周波数変化を捉えることにより、音声知覚を行なっていると考えられる。スペクトルピークの周波数変化は音声知覚において重要な手がかりであることから、本モデルを計算機音声認識に対して応

用できれば高精度化が可能となるだろう。

## 第 3 章

# スペクトルピーク追跡モデル

序論の本研究の特色にあるように 1 次聴覚野 (A1) におけるスペクトル知覚モデルを用いて、スペクトルピークを振幅、周波数、帯域幅、対称性の 4 パラメータで表現する。その 4 つのパラメータ (周波数、振幅、帯域幅、対称性) の時間変化を 2 次系システムを用いて予測し、任意の時刻における複数ピークの追跡を行なう。予測追跡された 4 パラメータを用いてスペクトルを再構成する。

本モデルは、入力スペクトル → 1 次聴覚野 (A1) モデル → 2 次系予測 → 追跡 → 再構成の 5 つのセクションから成り立っている (図 3.1)。以下、順に解説していく。

### 3.1 入力スペクトル

河本のモデルでは、入力であるスペクトル包絡を改良ケプストラム法 [13] により求めていた。しかし、この方法で求められたスペクトル包絡には、大きな正のバイアスがかかっていること、微細構造の変動の影響を受けてしまうことなどの欠点があった。このためスペクトルピーク周辺以外の周波数帯域におけるばらつきを抑えるために、前後 7 点による移動平均処理を行なっていた。

一方、不偏推定法 [14] は、対数スペクトルの推定値にバイアスが生じないように非線形平滑化を行なうことにより、改良ケプストラム法から求められるスペクトル包絡よりも微細構造を表現するスペクトル包絡を求めることができる。不偏推定法を用いることにより、移動平均処理を行なわなくてもフレームごとのはらつきを押えることができる。

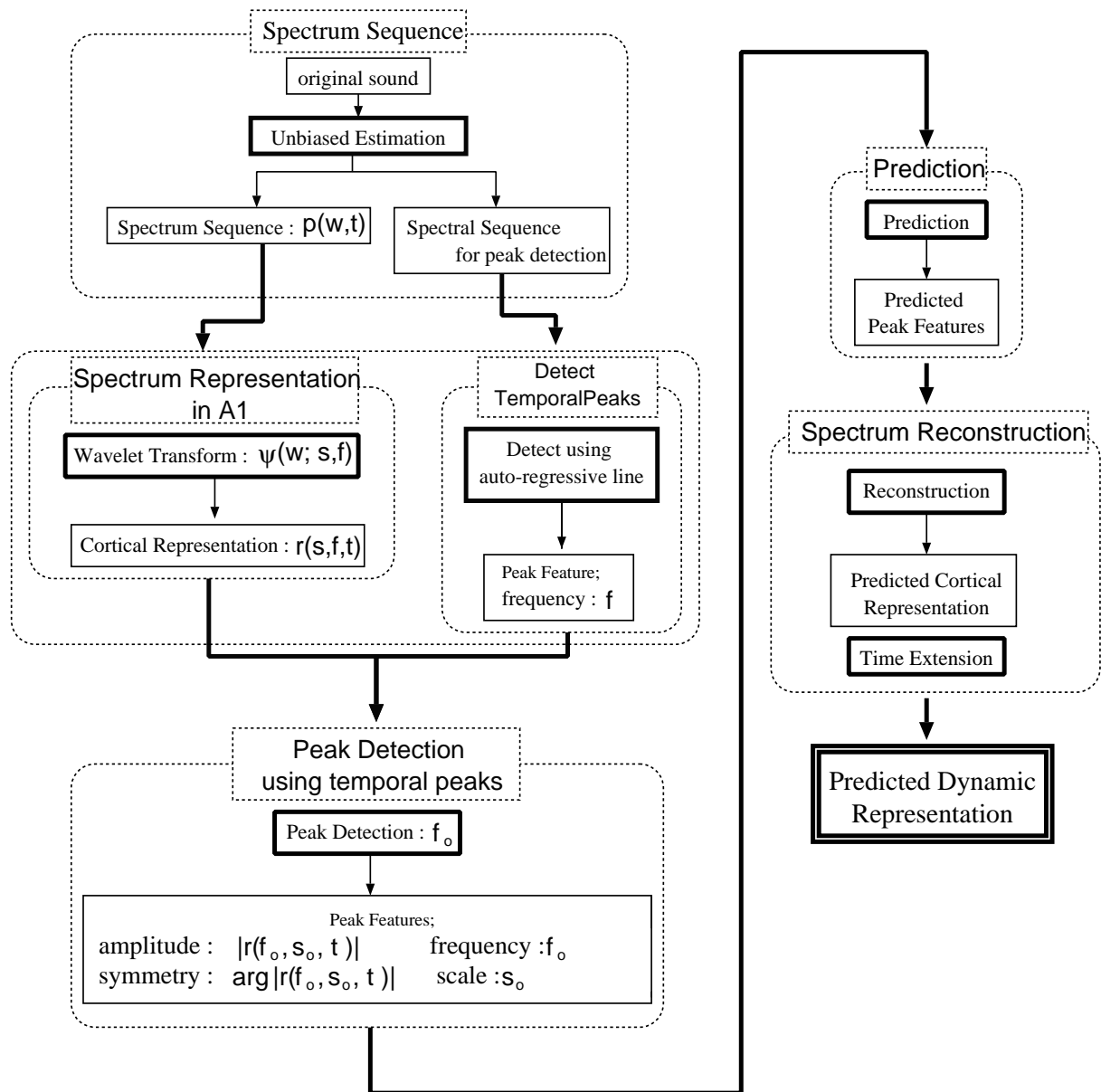


図 3.1: スペクトルピーク予測追跡モデルのフローチャート

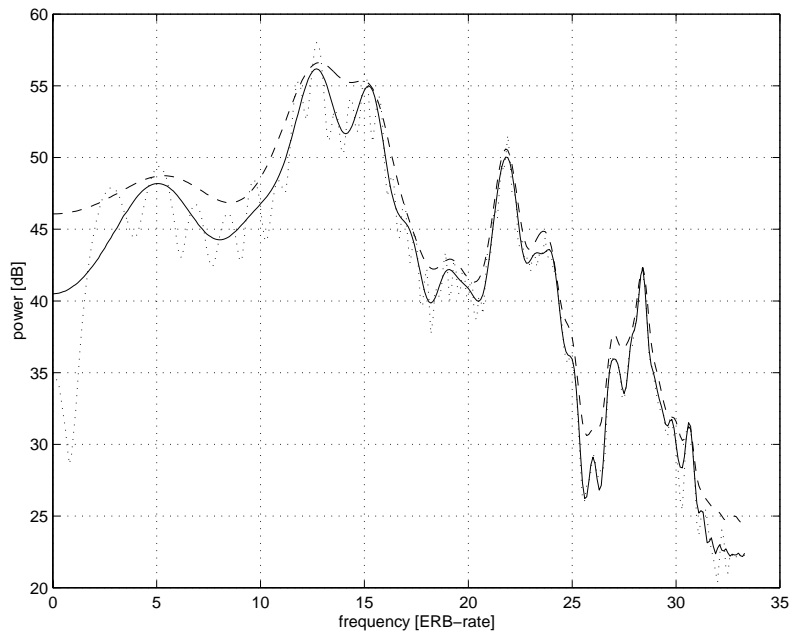


図 3.2: 両方法から得られるスペクトル ( 実線 : 不偏推定法 (unbias)、破線 : 改良ケプストラム法 (impcep)、点線 : FFT スペクトル )

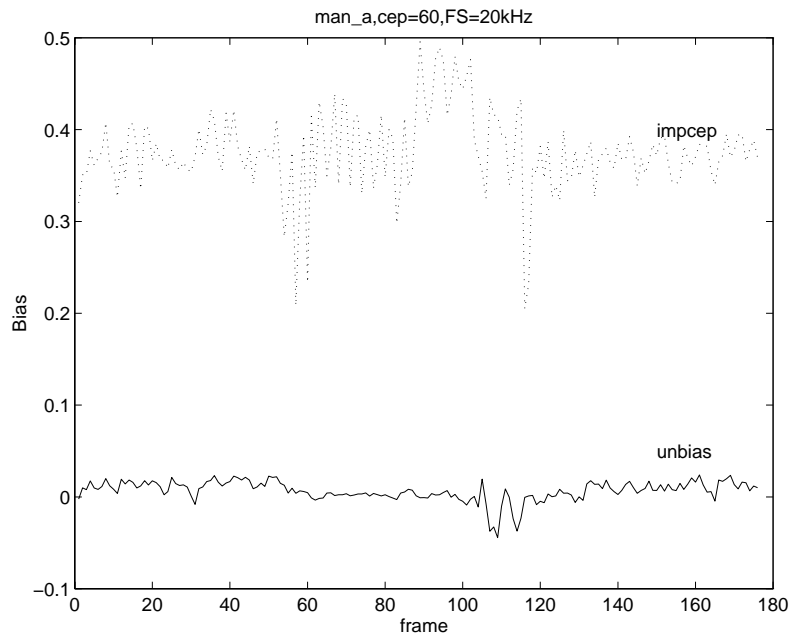


図 3.3: バイアスの比較 ( 実線 : 不偏推定法 (unbias)、点線 : 改良ケプストラム法 (impcep) )

そこで、本モデルでは、これらの欠点を解消するため不偏推定法を用い、音響信号よりスペクトル包絡を求めた。なお、ケプストラム次数は 60 次、サンプリング周波数は 20 kHz、フレーム長 25.6 msec、フレームシフト幅 6.4 msec とした。周波数は、基底膜振動など聴覚末梢系における生理学的及び心理物理学的知見と対応関係の良い ERB-rate[15] に変換する (式 (3.2))。

$$\text{ERB} = 24.7(4.37 * f[\text{kHz}] + 1) \quad (3.1)$$

$$\text{ERB} - \text{rate} = 21.4 \log_{10}(4.37 * f[\text{kHz}] + 1) \quad (3.2)$$

図 3.2、3.3 に両方法を比較した結果を示す。点線が ATR データベースの男性/a/(mau1001.ad) の原音声のスペクトル、破線が改良ケプストラム法より求めたスペクトル包絡、実線が不偏推定法より求めたスペクトル包絡である。図 3.4 には、不偏推定法から求めた本モデルの入力スペクトルを示す。前後 7 点による移動平均を施さなくても、改良ケプストラム法の問題点であった時間方向へのスペクトルのばたつきも押えられている。また、スペクトル微細構造を良く表現しており、より正確なピーク情報を得られることから改良ケプストラム法よりも、不偏推定法の方が本モデルには適していると言える。

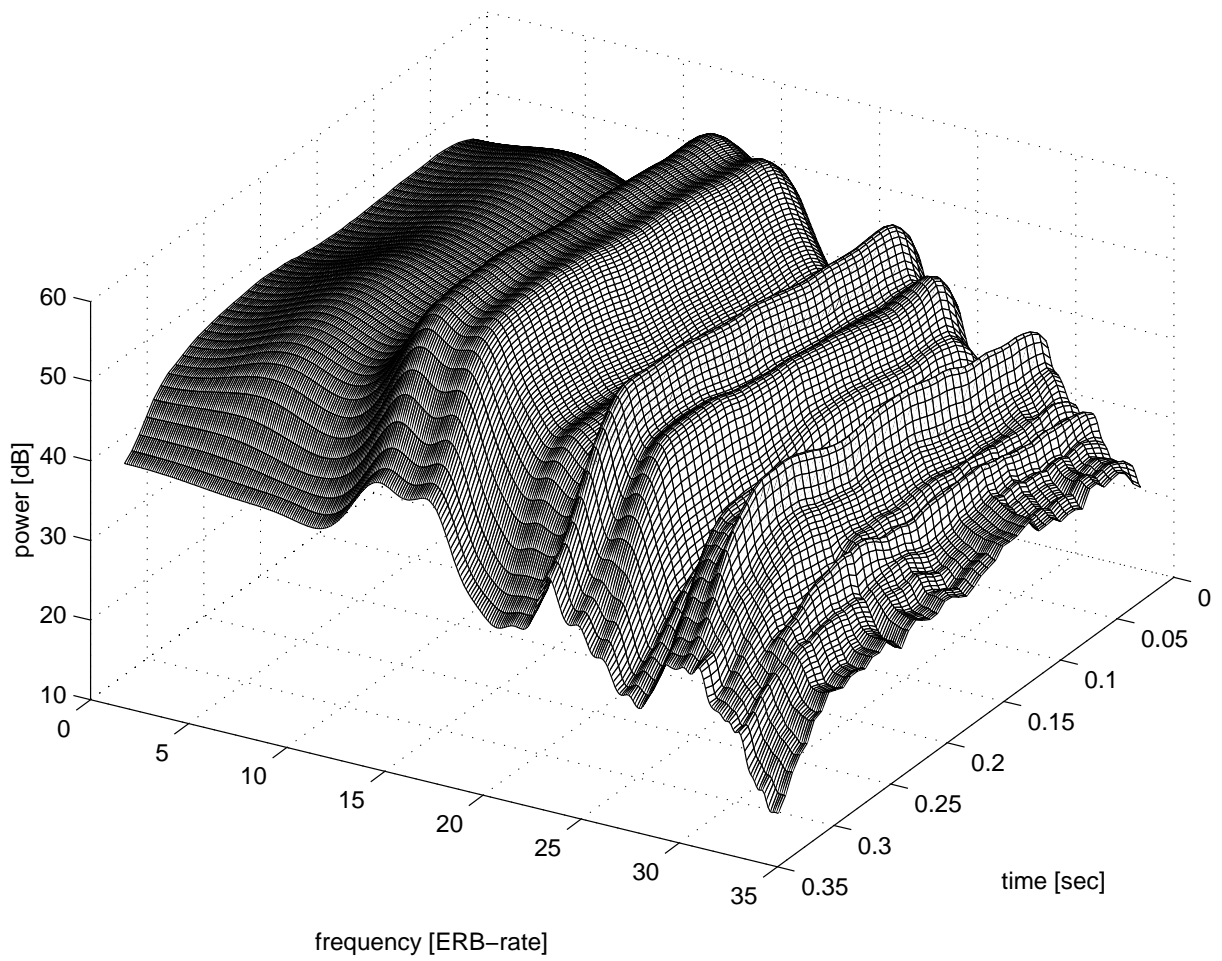


図 3.4: 入力スペクトル (男性: /あ/)



## 3.2 1次聴覚野 (A1) におけるスペクトル表現

Shamma らは、1次聴覚野応答領域の構造及び機能、聴覚末梢や聴神経の信号機能に関する生理学的知見を考慮して機能モデルを構築した。1次聴覚野においては特徴周波数が等しい細胞が集まり構成されている。各神経細胞群は特定の帯域幅と対称性を持っており、ある周波数に局在した周波数応答特性を持つことから、その応答特性を振幅、周波数、帯域幅、対称性の4パラメータで表現することができる。

Shamma らによって提案された、1次聴覚野 (A1) におけるスペクトル表現を実現するために、A1 における神経細胞群の応答を Gabor 関数  $\psi(\omega; s, f)$  で近似し、これを基本 wavelet として入力対数振幅スペクトル  $p(\omega, t)$  に対する周波数方向への wavelet 変換を行なう。以下に詳しく説明する。

### 3.2.1 周波数応答特性の表現

1次聴覚野 (A1) における神経細胞群の応答特性に着目し、側抑制の表現を考慮した神経細胞群の周波数応答関数を式 (3.3) で表される Gabor 関数を用いて表現する [11]。

$$\psi(\omega) = \exp\left(-\left(3\frac{10^{-1/8}-1}{1+10^{-1/8}}\phi_c\omega\right)^2\right)\exp(i\phi_c\omega) \quad (3.3)$$

ここで、 $\phi_c$  は Gabor 関数の中心角周波数を表す。この周波数応答関数におけるピークの周波数、帯域幅、対称性を変化させることにより、任意の神経細胞群の周波数応答特性を表現することができる。

### 3.2.2 A1 における神経細胞群の応答

1次聴覚野 (A1) におけるスペクトル表現では、各神経細胞群の応答は入力スペクトルと応答特性とのマッチングの度合であり、入力スペクトルと応答特性の内積を式 (3.4) のように定義する。 $\psi(\omega; s, f)$  を基本 wavelet とした入力スペクトル  $p(\omega, t)$  に対する周波数方向への wavelet 変換 [16][17] として表すことができる。以上のことから、時刻  $t$  における神経細胞群の応答  $r(s, f, t)$  を式 (3.4) で記述する。

$$\begin{aligned} r(s, f, t) &= \langle p(\omega, t), \psi(\omega; s, f) \rangle_\omega \\ &= \frac{1}{|s|^{1/2}} \int_{-\infty}^{\infty} \psi^*\left(\frac{\omega-f}{s}\right) p(\omega, t) d\omega \end{aligned} \quad (3.4)$$

ここで、 $s$ 、 $f$ 、 $t$  はそれぞれ帯域幅、A1 での周波数位置、時間を、 $|r(s, f, t)|$  は振幅、位相項  $\arg |r(s, f, t)|$  は対称性を表す。また、 $*$  は複素共役を表す。

### 3.3 スペクトルピーク抽出アルゴリズム

純音やスィープ音のように単一ピークしか持たない信号を対象にした場合、ピーク周波数の抽出は単純に振幅が最大値となる値をピーク周波数とすればよく、大変容易であった。しかし、音声などのホルマント (formant) のように複数個のピークが存在するような場合、ピーク周波数の抽出は容易ではない。

不偏推定法を用いて得られるスペクトル包絡は、ケプストラム次数が大きければ大きいほどより微細構造を表現するスペクトル包絡を得ることができる。しかし、ピーク以外の微小変動をもピークとして抽出してしまう問題がある。そこで、本モデルでは図 3.5 のように、ケプストラム次数を下げ得られるスペクトル包絡に対して、フレームごとに自己回帰直線を引く。この直線よりも大きな値で、かつ前後 3 点で比較し両隣の値よりも大きな値をスペクトルのピーク周波数として抽出する。この手法により、良い精度で複数のピーク周波数を得ることができる。図 3.5 は合成母音 /a/ から、50 次までのケプストラム次数を用いて得たスペクトル包絡から、自己回帰直線を用いてピークを抽出している。

そして、この手法で得られたピーク周波数を式 (3.4) で振幅  $|r(s, f, t)|$  の極大値を与えるピーク周波数  $f$  の候補とする。図 3.6 のように、候補  $f$  の近傍で  $|r(s, f, t)|$  の極大値を探索し、極大値となる  $|r(s_0, f_0, t)|$  から帯域幅  $s_0(t)$ 、ピーク周波数  $f_0(t)$ 、ピーク振幅レベル  $|r(s_0, f_0, t)|$ 、対称性  $\arg |r(s_0, f_0, t)|$  を決定する。

### 3.4 パラメータの予測法

本モデルでは、式 (3.5) のような 2 階微分方程式で表されるシステムによって、スペクトルピークを表す 4 パラメータ各々についての予測を行なう。

$$a(1-w)y''(t) + \{b(1-w) + cw\}y'(t) + y(t) = (1-w)x(t) \quad (3.5)$$

ここで、 $a$ 、 $b$ 、 $c$  は定数であり  $x(t)$ 、 $y(t)$  は 4 パラメータに対応する入力、出力、 $w$  は重み関数を表す。

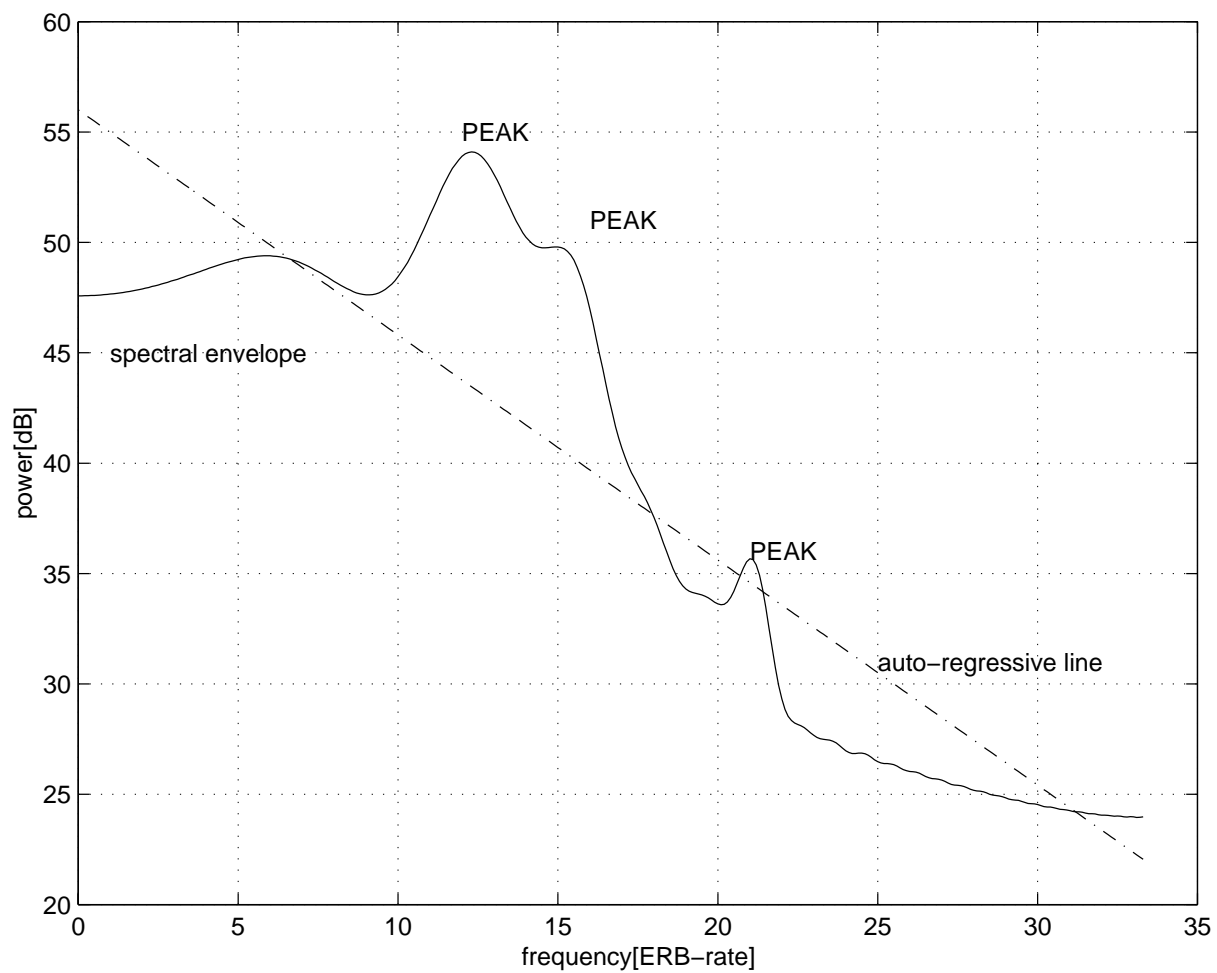


図 3.5: 自己回帰直線を用いたピーク抽出 (実線: スペクトル包絡、鎖線: 自己回帰直線)

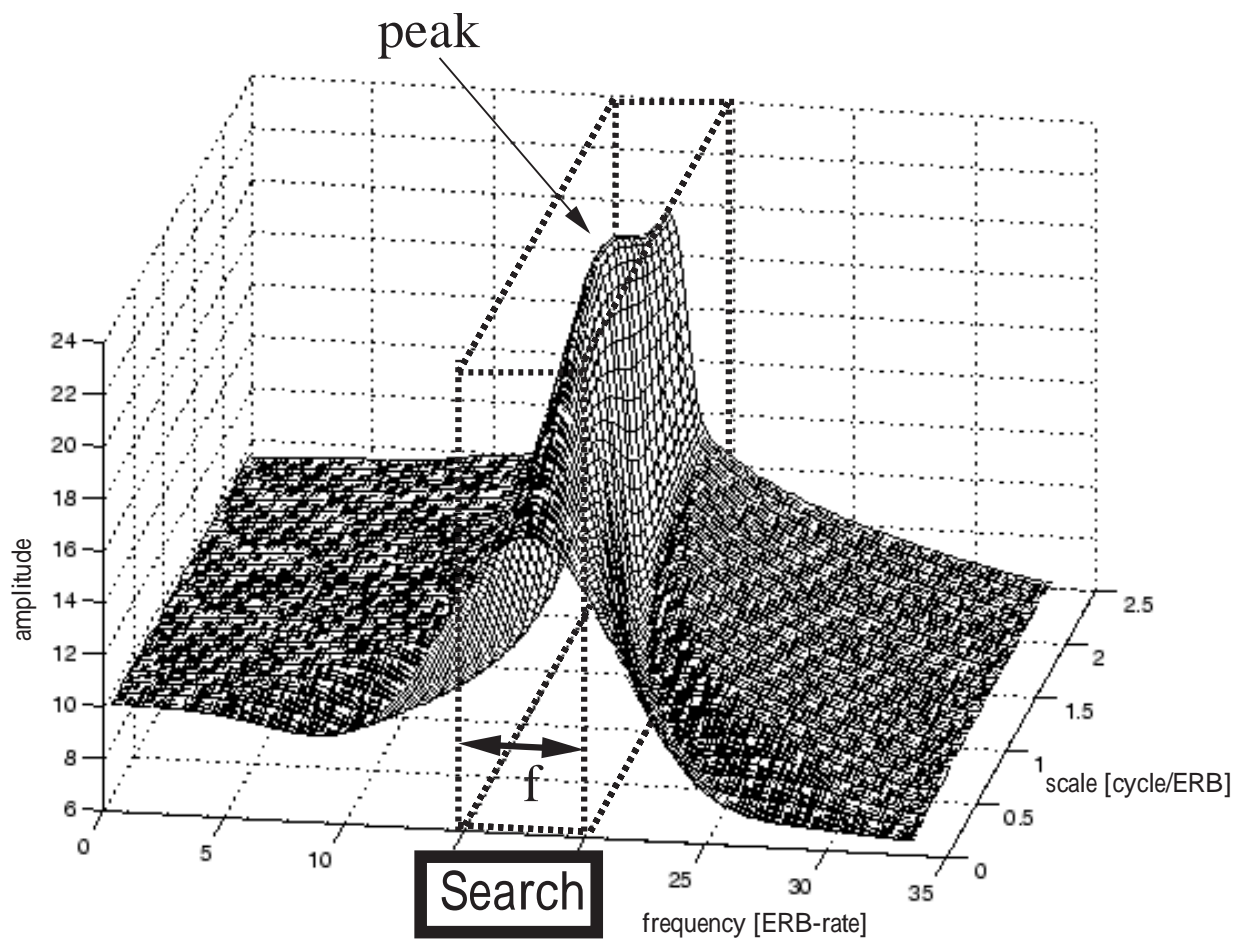


図 3.6: A1 表現からのピーク抽出

### 3.4.1 雑音区間の抽出

一般に全く雑音のないクリーンな状況というのは考えられない。そこで、雑音の存在する場合を考える。入力信号のあるピークの帯域幅に存在する雑音の振幅レベルが、その信号の振幅レベルよりも大きくなった時、雑音によって目的信号が置換されたと判断する<sup>1</sup>。

本モデルでは信号の振幅レベルが大きく変化しないものを仮定する。

フレームごとに入力スペクトルの振幅から実効値を求め、1フレーム前の値と比較する。急激に変化量が大きくなった時、雑音の開始時点とし、終了時点は入力信号を用いて得られるある閾値よりも実効値が小さくなった時点とした。例として、1464 Hz ~ 2297 Hz へ 1000 msec で上昇するスイープ音に、500 msec を中心として、スイープ音よりも大きい振幅レベルを持つ、150 msec の白色雑音を付加する。同様に、2297 Hz ~ 1464 Hz へ 1 sec で下降するスイープ音に、1500 msec を中心として、スイープ音の振幅レベルよりも大きい振幅レベルを持つ、150 msec のホワイトノイズを付加する。これを1周期とする信号の雑音判断を図 3.7に示す。

### 3.4.2 通常区間での予測機構

通常区間では、式 (3.5) に中心差分演算を行なうことによって式 (3.9) のように変形する。

中心差分演算では

$$y(t)' = \frac{y[n+1] - y[n-1]}{2} \quad (3.6)$$

$$y(t)'' = y[n+1] - 2y[n] + y[n-1] \quad (3.7)$$

として式 (3.5) に代入すると

$$\begin{aligned} \{2a(1-w) + b(1-w) + cw\}y[n+1] = \\ 2(1-w)x[n] - \{2 - 4a(1-w)\}y[n] - \{2a(1-w) - b(1-w) - cw\}y[n-1] \end{aligned} \quad (3.8)$$

式 (3.8) において  $n = n - 1$  とすると

$$y[n] = \frac{2(1-w)x[n-1] - \{2 - 4a(1-w)\}y[n-1] - \{2a(1-w) - b(1-w) - cw\}y[n-2]}{\{2a(1-w) + b(1-w) + cw\}} \quad (3.9)$$

<sup>1</sup>本モデルでは、雑音として白色雑音のみを用いている。また、重み関数は雑音区間では  $w = 0$  とし通常区間では  $w = 1$  とする

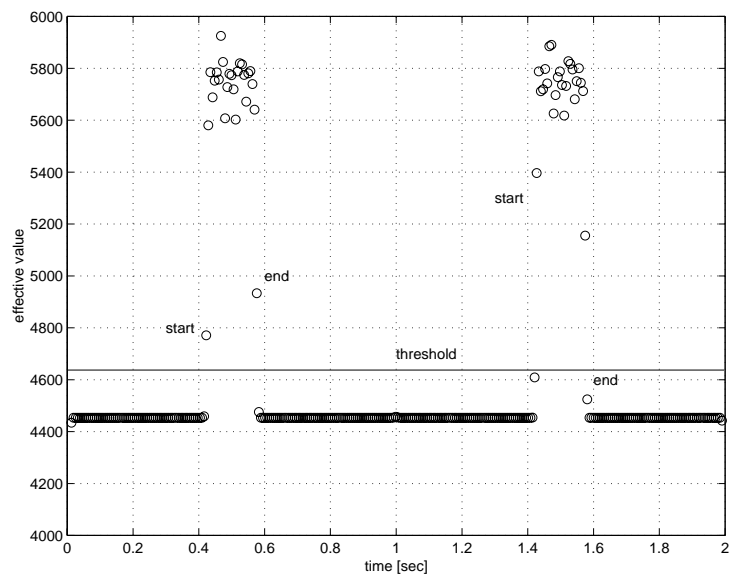
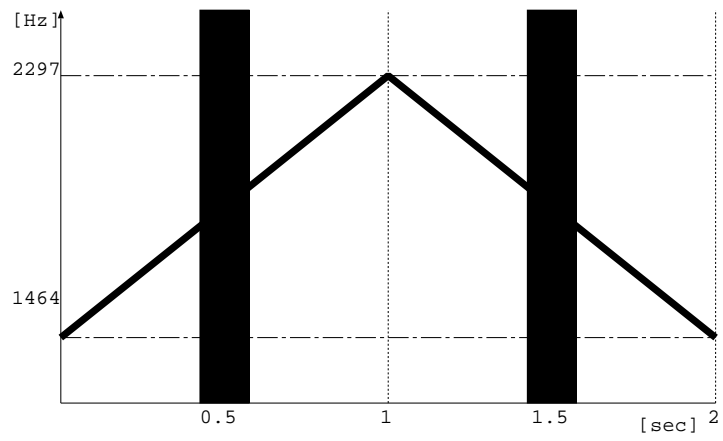


図 3.7: 実効値を用いた雑音の判断 ( O印 : 実効値、実線 : 閾値 )

の関係式を得る。ここで、

$$w = 0 \quad (3.10)$$

$$G = \frac{2}{2a + b} \quad (3.11)$$

$$\alpha_1 = \frac{2(1 - 2a)}{2a + b} \quad (3.12)$$

$$\alpha_2 = \frac{2a - b}{2a + b} \quad (3.13)$$

とすると

$$y[n] = Gx[n - 1] - \alpha_1 y[n - 1] - \alpha_2 y[n - 2] \quad (3.14)$$

となり、相川らに提案された2次系システムと同様の関係式を得る。式(3.14)のように、任意の時刻  $n$  における出力  $y[n]$  はその時刻の1時刻前の入力  $x[n - 1]$  と、1時刻および2時刻前の出力  $y[n - 1]$  と  $y[n - 2]$  の線形結合によって求めることができる。式(3.14)において、変化のない一定の入力に対して、入力と同じ出力が得られるように、

$$G = 1 + \alpha_1 + \alpha_2 \quad (3.15)$$

とする。

### 3.4.3 雑音区間での予測機構

倉片らは、心理物理実験から雑音が後続する周波数変化音の主観的な持続時間が、雑音が後続しない周波数変化音の主観的持続時間に比べ、過大評価され、また上昇(下降)スイープ音の終端に雑音を付加した場合の知覚結果は、雑音がない場合の時よりもより高く(低く)知覚されことから終端のピッチも変化方向を延長する側にシフトする傾向があることを示した。このことから、雑音内では、雑音前の音情報を保持しつつ、スペクトルピーク追跡を行なっていると思われる。

そこでスペクトルピークの周波数、対称性追跡には増(減)が一定である変化を記述するために式(3.16)を用いて追跡を行なう。

$$y[n] = y[n - 1] + (y[n - 1] - y[n - 2]) \quad (3.16)$$

入力を停止するということは、式(3.14)において入力  $x$  をゼロにすることになる。その場合、追跡周波数は急激にゼロに向かって移動する。しかし、このような追跡は知覚的

連続性、予測効果を考慮すると考えにくい。相川らは文献 [9] の中で、入力が停止すると次第にピークは広がり、ピーク周波数がはっきりしなくなると述べている。そこで、本モデルではスペクトルピークの振幅の追跡には後退差分演算をすることによって式 (3.5) を式 (3.20) のように変形した式を用いて追跡を行なう。後退差分演算では

$$y(t)' = y[n] - y[n - 1] \quad (3.17)$$

$$y(t)'' = y[n] - 2y[n - 1] + y[n - 2] \quad (3.18)$$

として式 (3.5) に代入すると

$$y[n] = \frac{(1 - w)x[n] + \{2a(1 - w) + b(1 - w) + cw\}y[n - 1] - a(1 - w)y[n - 2]}{\{a(1 - w) + b(1 - w) + cw + 1\}} \quad (3.19)$$

の関係式を得る。ここで  $w = 1$  とすると

$$y[n] = \frac{c}{c + 1}y[n - 1] \quad (3.20)$$

となる。また、帯域幅は

$$\text{amplitude} * \text{scale} = \text{const.} \quad (3.21)$$

の関係から求めた。

上記の雑音判断の例に用いた同じ信号で、スペクトルピークの 4 パラメータ各々について予測を行なった例を図 3.8、3.9 に示す。但し、通常区間では 2 次系のシステムとして固有周波数が 4 Hz、時定数が 0.0398 の臨界制動となるように  $a, b, c$  を設定した。雑音区間では  $c = 0.0064$  とした。これらの値は、計算機シミュレーションを行なった結果から経験的に求めた。

### 3.5 複数ピークの追跡

複数ピークが存在する場合、式 (3.14) における、次入力  $x[n - 1]$  をどのピークの次入力  $x[n - 1]$  から選ぶかが問題となる。

Bregman は聴覚情景解析 (Auditory Scene Analysis)[1] の中で、次の 4 つの発見的規則を枚挙して聴覚の初期段階における音の分擬やグループ化について説明している。

1. 変化は急激には起こらない。



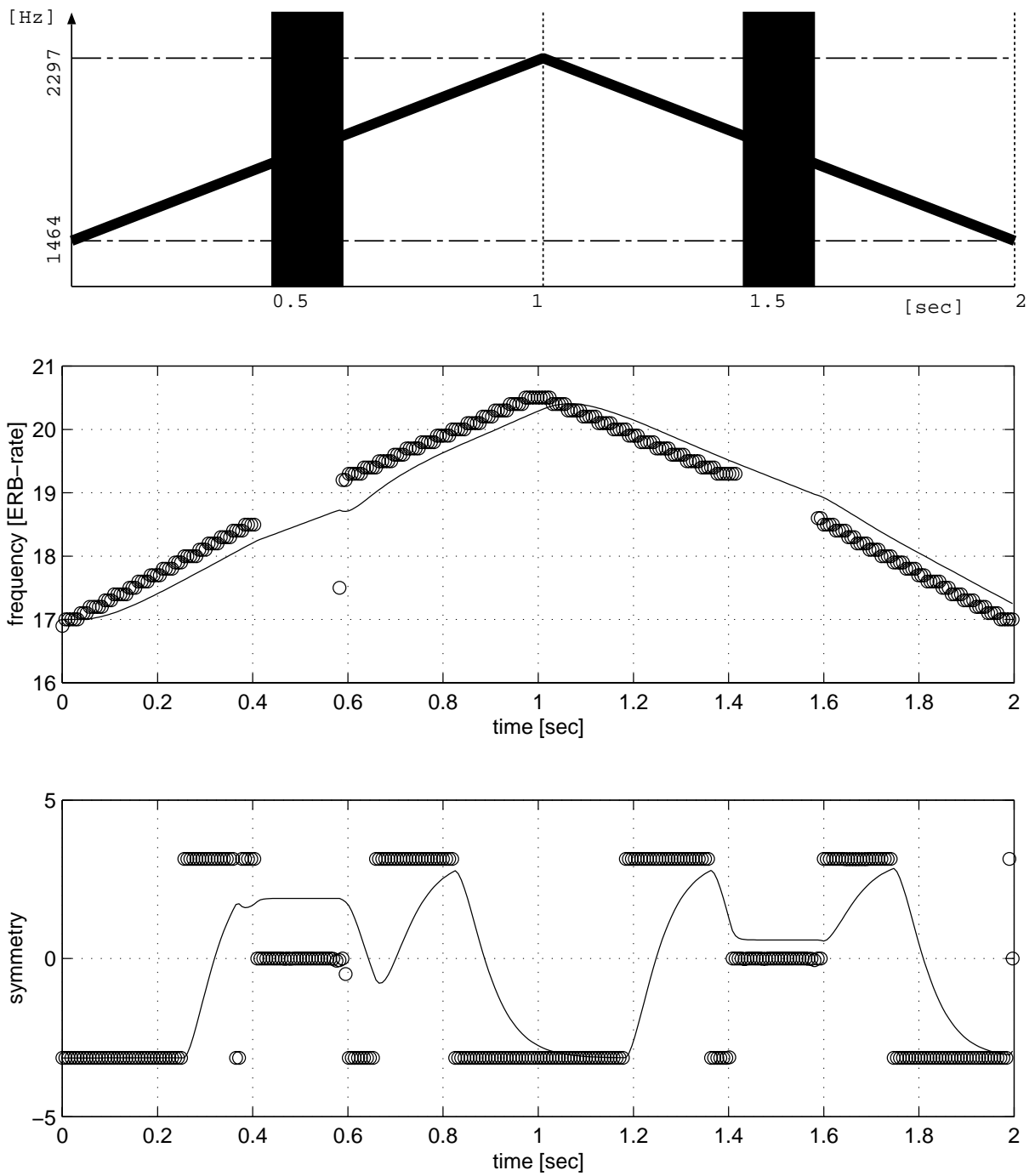


図 3.8: パラメータの予測結果 : シミュレーションデータの概図 [上段]、ピークの周波数 [中段]、対称性の予測 [下段] (実線 : 予測追跡結果、○印 : 抽出されたピーク情報)

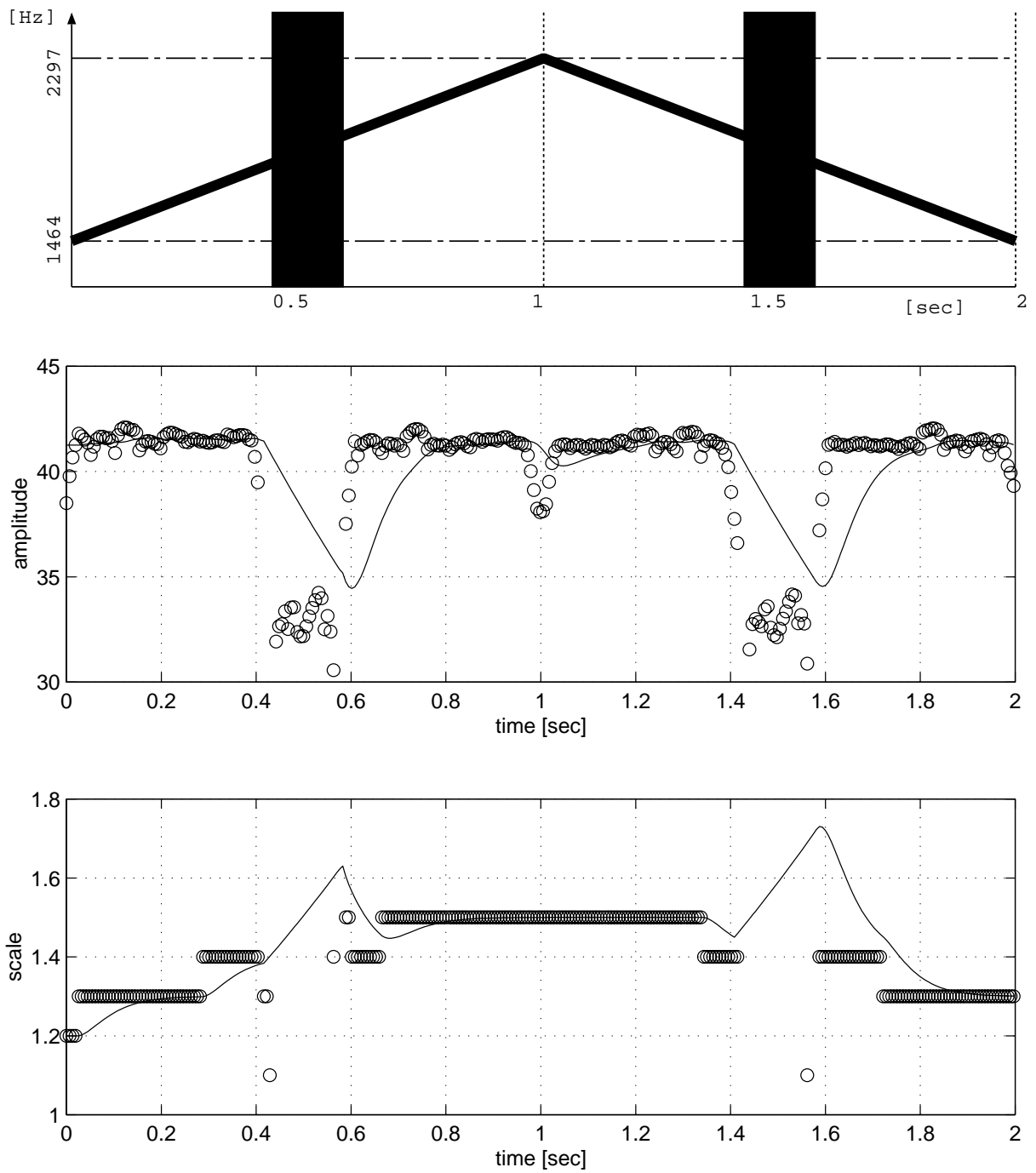


図 3.9: パラメータの予測結果: シミュレーションデータの概図 [上段]、ピークの振幅 [中段]、帯域幅の予測 [下段] (実線: 予測追跡結果、○印: 抽出されたピーク情報)

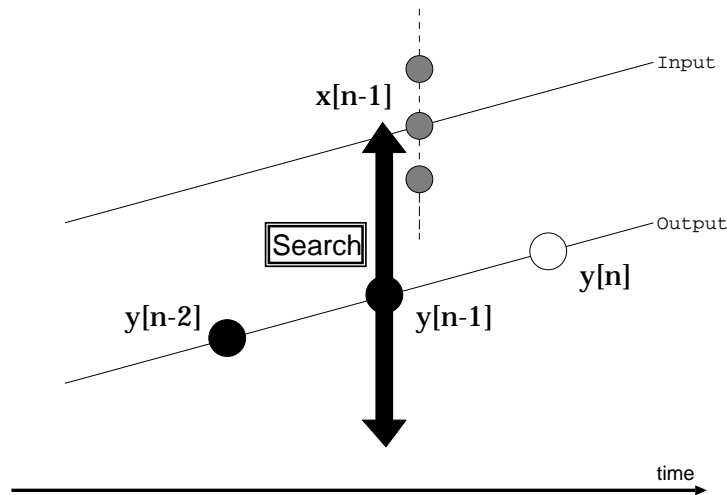


図 3.10: 複数ピークの追跡

2. 物が繰り返し振動する時、共通の基本周波数の整数倍の成分が発生する。
3. 関係のない音は同時に始まったり終わったりはしない。
4. 一つの音響事象に生じる変化は、その音を構成する成分に同時に同じような影響を与える。

そこで、本モデルでは、1. の音脈の形成に関わる規則である「よい連続の原理や周波数近傍の変理の概念」を追跡系に導入することによって複数ピークの追跡を実現する。具体的には、図 3.10 のように式 (3.14) で表される追跡系において、1 時刻前の出力  $y[n-1]$  から 1 時刻前の入力  $x[n-1]$  を選定する。その時、ある出力  $y[n-1]$  の近傍に複数の入力  $x[n-1]$  が存在する場合は、Bregman の規則 1 に基づき探索範囲を定め、出力  $y[n-1]$  に最も近い  $x[n-1]$  を次の入力とする。

本モデルの予測・追跡機構を用いた例として、1 kHz ~ 3 kHz へ上昇するスイープ音と 3 kHz ~ 1 kHz へ下降するスイープ音の 2 つの信号が存在する場合を固有周波数 4 Hz、減衰定数 0.0398 として予測・追跡した結果を図 3.11 に示す。

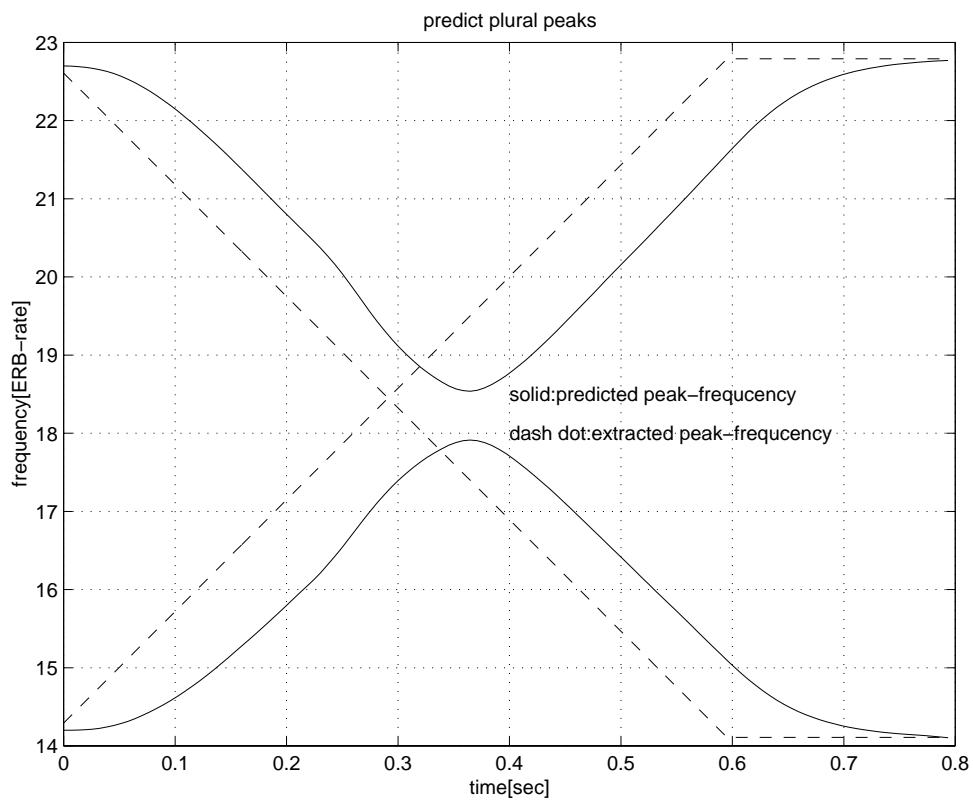


図 3.11: 本モデルを用いた複数ピークの予測・追跡の例 ( 実線 : 予測追跡結果、破線 : 原信号のピーク周波数 )

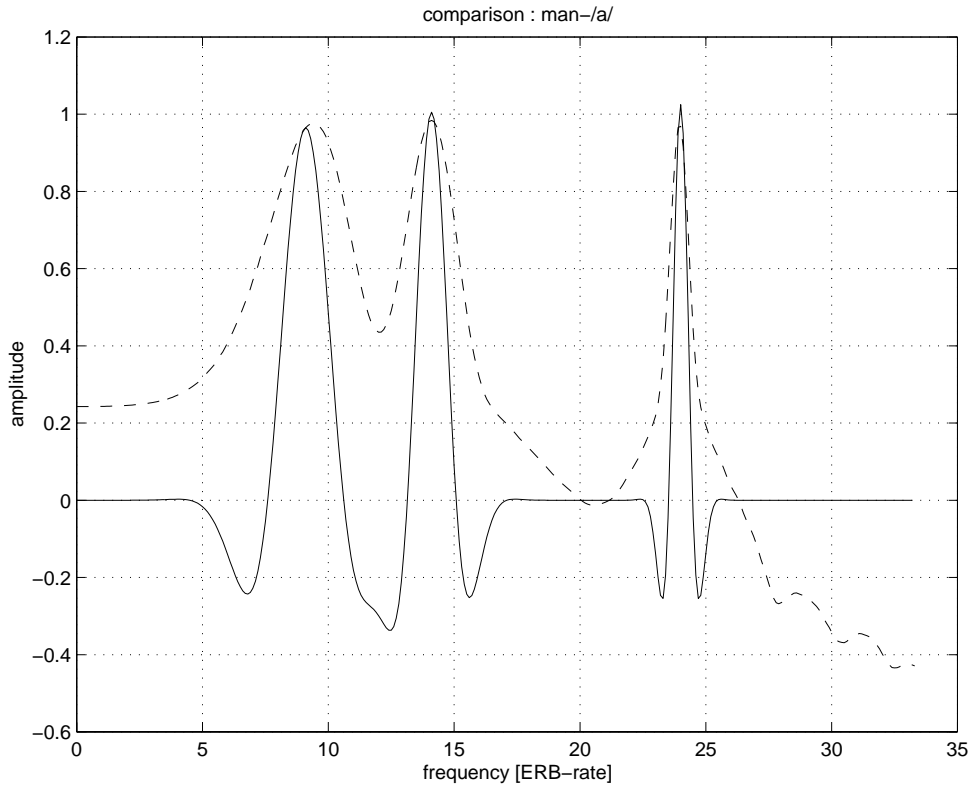


図 3.12: スペクトル再構成結果 ( 実線 : 再構成スペクトル、破線 : 原信号スペクトル )

### 3.6 4 パラメータによるスペクトルの再構成

式 (3.4) で、各ピークごとに予測・追跡された周波数  $\hat{f}_0$ 、振幅  $|\hat{r}(\hat{s}_0, \hat{f}_0, t)|$ 、帯域幅  $\hat{s}_0$ 、対称性  $\arg \hat{r}(\hat{s}_0, \hat{f}_0, t)$  を用いて、式 (3.22) で表される、逆 wavelet 変換を行ないスペクトルを再構成する。本モデルでは、予測追跡に関してスペクトルピークの 4 パラメータで表現される情報を扱っている。そのピーク情報のみを用いた場合においても、図 3.12 のようにピーク及びその付近の周波数帯域におけるスペクトルの概形を表現することが可能である。

$$p(\omega, t) = \frac{1}{2\pi C_\phi} \int_{-\infty}^{\infty} \frac{1}{|\hat{s}_0|^{\frac{1}{2}}} \hat{r}(\hat{s}_0, \hat{f}_0, t) \psi\left(\frac{\omega - \hat{f}_0}{\hat{s}_0}\right) \frac{d\hat{s}_0 d\hat{f}_0}{\hat{s}_0^2} \quad (3.22)$$

## 第 4 章

# 周波数変化音予測・追跡シミュレーション

3 章で説明した本スペクトルピーク予測追跡モデルの妥当性を検証する目的で計算機シミュレーションを行なった。以下の節において、複数の対数線形スイープ音、クロスする 2 本の対数線形スイープ音、雑音による中断のある対数線形スイープ音、合成音に対するシミュレーションについて述べる。なお、シミュレーションの条件としては表 4.1 の設定値を使用した。

### 4.1 複数の対数線形スイープ音に対するシミュレーション

#### 4.1.1 目的

河本のモデルでは、予測追跡が不可能だった複数のスペクトルピークの同時追跡が可能となったこと、固有周波数と減衰定数を制御することにより、相川らに報告されているリンギング効果を模擬できることを示す。

#### 4.1.2 シミュレーションに使用するデータ

600 msec で 0.5 kHz ~ 1 kHz、1 kHz ~ 2 kHz へ上昇する対数線形スイープ音、及び 3 kHz ~ 2.5 kHz へ下降する 3 本の対数線形スイープ音の各々に 200 msec の trailer を付加したのちから成る合成音 (図 4.1) と、0 ~ 200 msec に 1 kHz ~ 2 kHz へ上昇、200 ~ 400 msec は 2 kHz の一定、400 ~ 600 msec に 2 kHz ~ 3 kHz へ上昇する対数線形スイープ音、0 ~ 200 msec に 4 kHz ~ 3 kHz へ下降、200 ~ 400 msec は 3 kHz の一定、400 ~ 600 msec に 3

表 4.1: シミュレーション条件

パラメータ	設定値
サンプリング周波数	20 [kHz]
フレーム長	25.6 [msec]
フレーム周期	6.4 [msec]
窓関数	hamming 窓
FFT ポイント	8192
チャンネル数	32
中心周波数	0.5 [cycle/ERB]

kHz ~ 4 kHz へ再び上昇する対数線形スイープ音 (図 4.2) を、ケプストラム次数 30 次で不偏推定法より求めたスペクトル包絡を入力信号とした。但し、信号の立ち上がり、立ち下がりには 5 msec のテーパーをかけた。

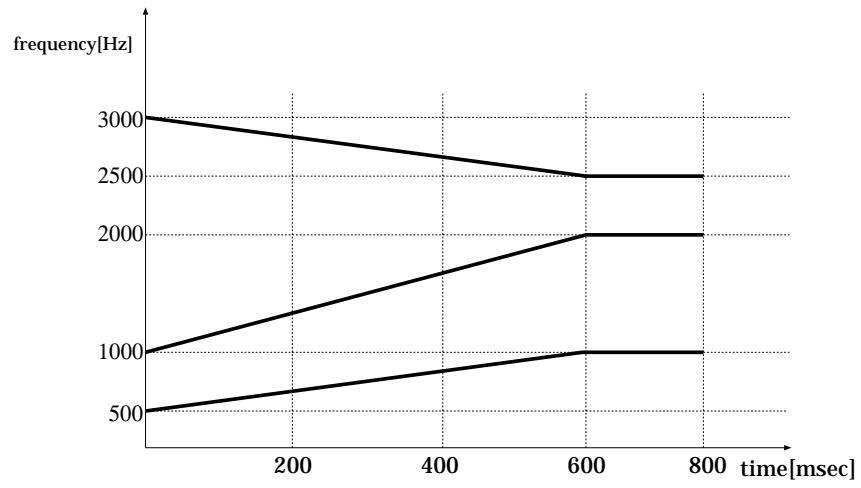


図 4.1: シミュレーションデータ 1

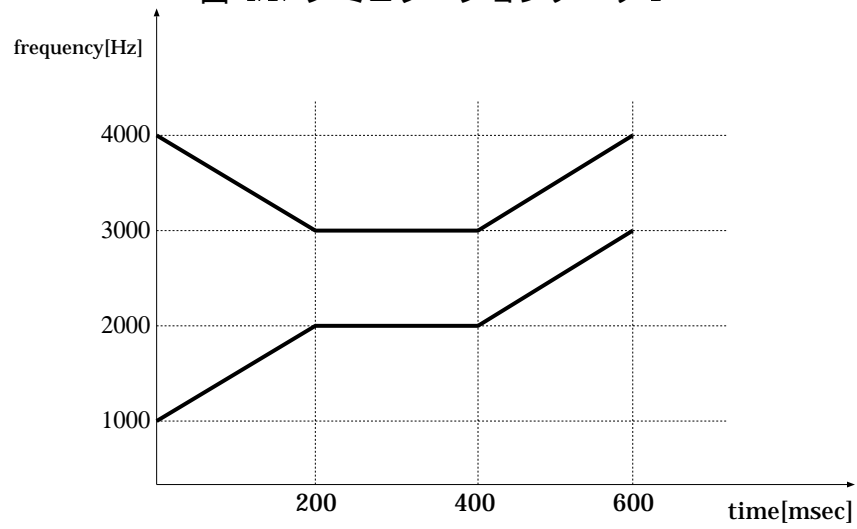


図 4.2: シミュレーションデータ 2



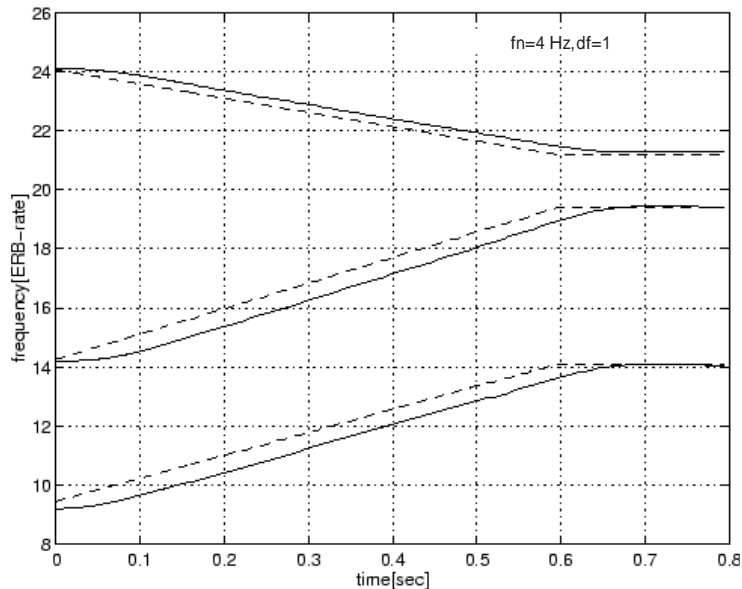


図 4.3: ピーク周波数の予測・追跡結果 ( 実線 : 予測追跡結果、破線 : 原信号のピーク周波数 )

#### 4.1.3 シミュレーション結果

図 4.1の合成音を固有周波数を 4 Hz, 減衰定数を 1 として本モデルで予測・追跡した結果を図 4.3、図 4.4 に示す。複数個のピークを持つ、上昇・下降する周波数変化音を予測・追跡することが可能となった。

図 4.5のように固有周波数と減衰定数を制御することで、停留部でのピッチの揺らぎ(リングング)を再現できる。実線は固有周波数 5 Hz、減衰定数 1、斜線は固有周波数 8 Hz、減衰定数 0.63 とした。また、図 4.5には、本モデルで固有周波数 5 Hz、減衰定数 1 として予測・追跡を行なった結果を示す。

#### 4.1.4 考察

文献 [9] で相川らが行なったものと同様のシミュレーションを行なった。ピーク周波数のみでなく、スペクトル構造を保持しつつ、予測追跡が行なわれていることがわかる。これは河本のモデルでは不可能なことであった。また、聴覚の過渡応答によって引き起こされるリングングもパラメータを制御することで模擬可能であることが分かった。

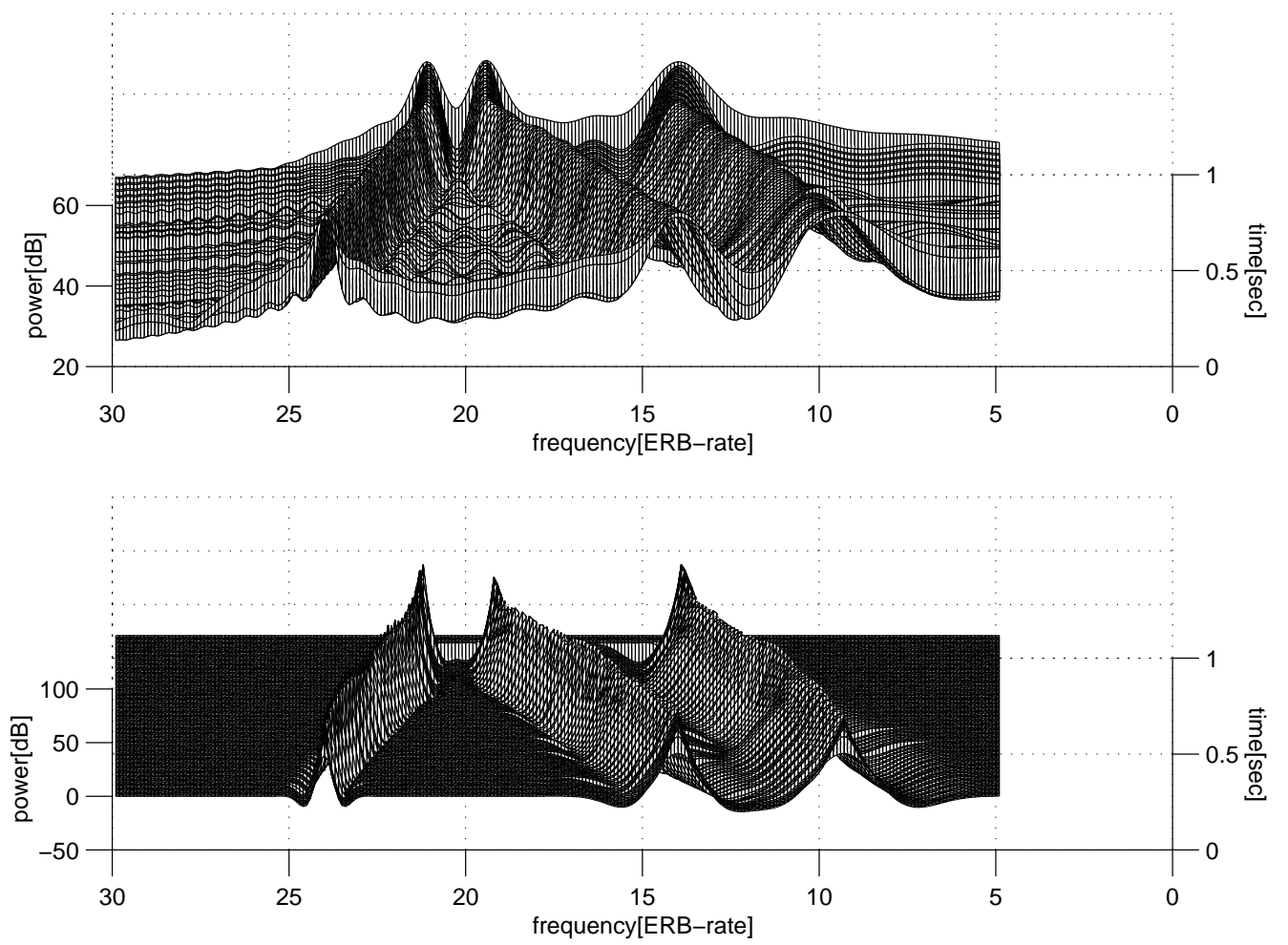


図 4.4: 入力スペクトル [上]、予測追跡されたスペクトル [下]

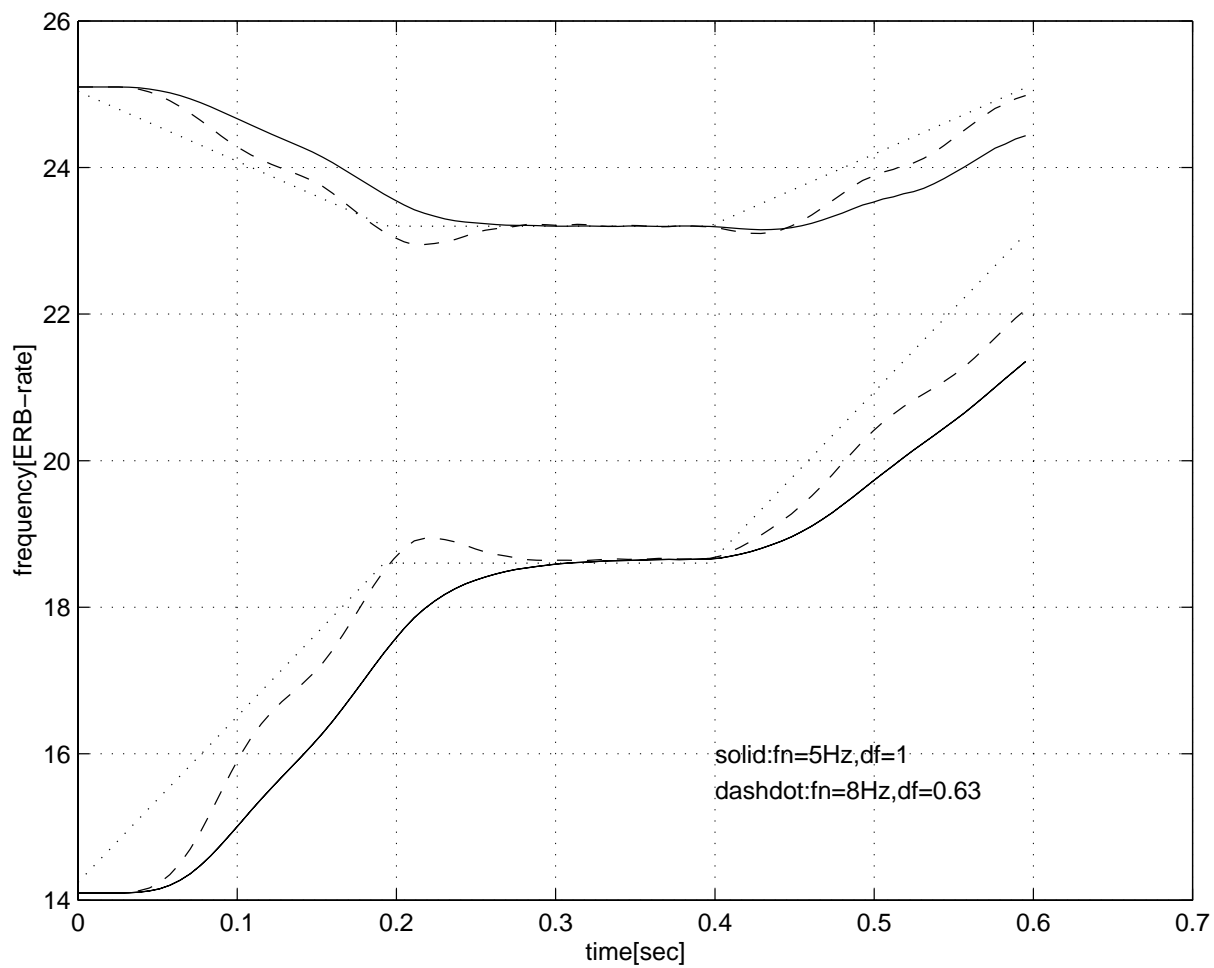


図 4.5: リンギングの制御 ( 実線 :  $fn=5\text{ Hz}, df=1$ 、破線 :  $fn=8\text{ Hz}, df=0.63$ 、点線 : 原信号のピーク周波数 )

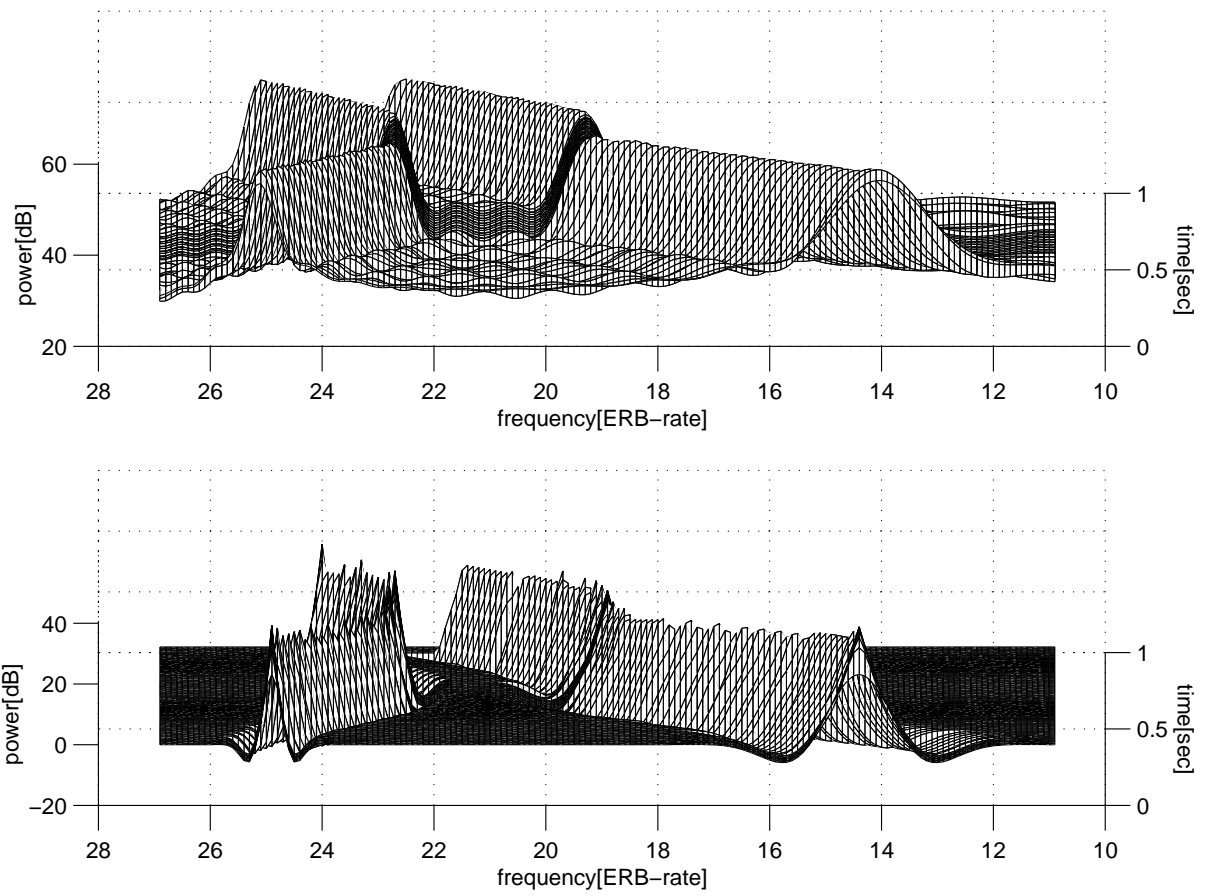


図 4.6: 入力スペクトル [上]、予測追跡されたスペクトル [下]

## 4.2 2本のクロスする対数線形スイープ音に対するシミュレーション

### 4.2.1 目的

Bregman は聴覚情景解析の中で、周波数が高い方から低い方へ変化するスイープ音と、低い方から高い方へ変化するスイープ音からなる合成音の知覚は、図 4.7 のような X のパターンではなく、図 4.8 のような 2 つの V パターンに知覚されると報告している (バウンド効果)。本モデルで固有周波数と減衰定数を制御することにより、バウンド効果を模擬できることを示す。

### 4.2.2 シミュレーションに使用するデータ

図 4.7 に示すように、600 msec で 1 kHz ~ 3 kHz へ上昇、3 kHz ~ 1 kHz へ下降する対数線形スイープ音から成る合成音からケプストラム次数 30 次で不偏推定法より求めたスペクトル包絡を入力信号とした。但し、信号の立ち上がり、立ち下がりには 5 msec のテーパをかけた。

### 4.2.3 シミュレーション結果

図 4.9 の合成音を固有周波数を 4 Hz、減衰定数を 1 として本モデルで予測・追跡した結果を図 4.10 に示す。交差するスイープ音を固有周波数、減衰定数を制御することによって、バウンド効果を模擬することが可能となった。

### 4.2.4 考察

Bregman によるデモストレーション [18]、倉片らの報告と同様に上下逆方向へ変化する線形対数スイープ音が交差した場合の知覚現象を、パラメータを制御することによって模擬できた。

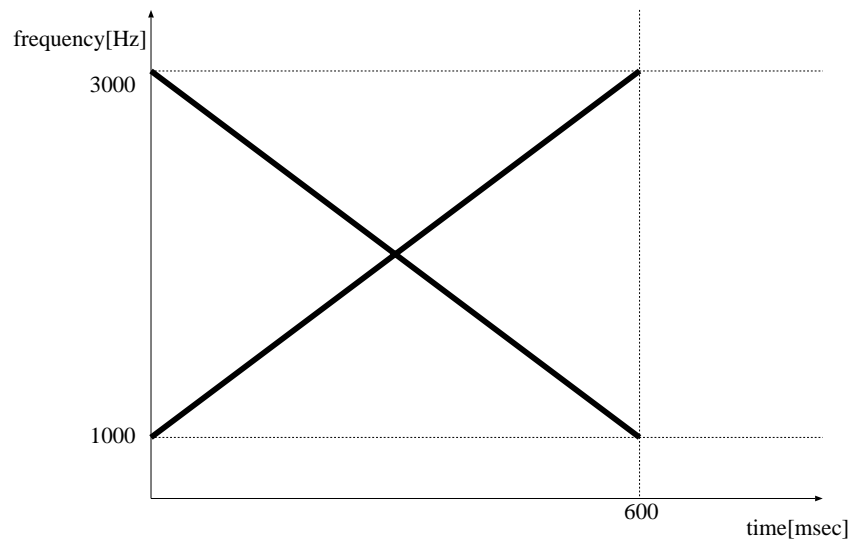


図 4.7: クロスするスイープ音 ( Xパターン )

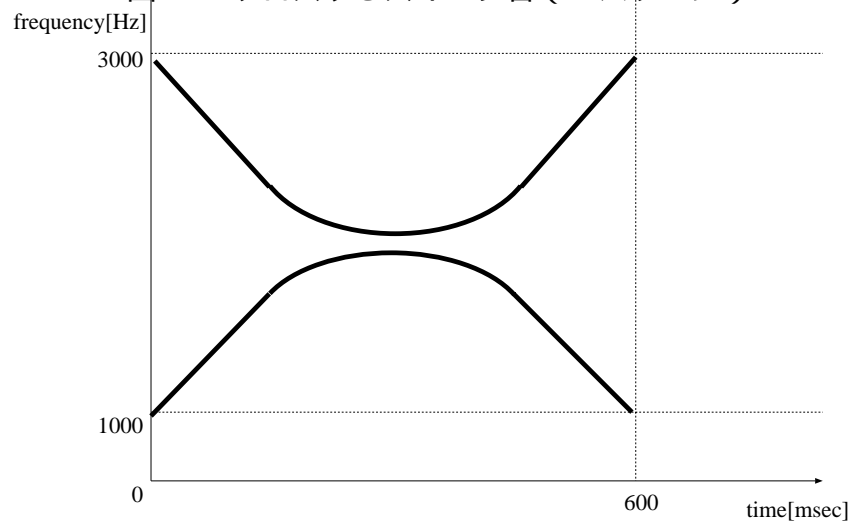


図 4.8: バウンド効果 ( Vパターン )

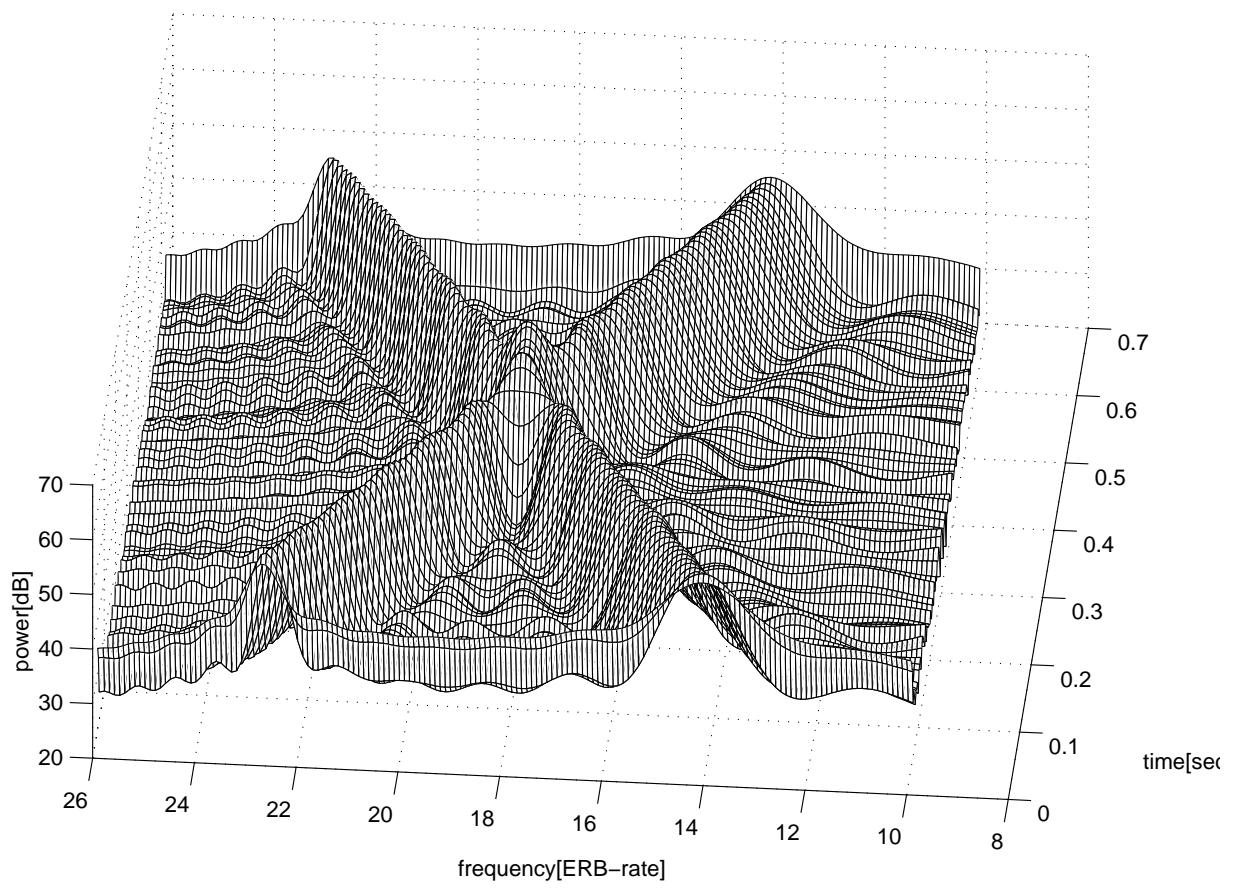


図 4.9: 交差するスイープ音の入カスペクトル

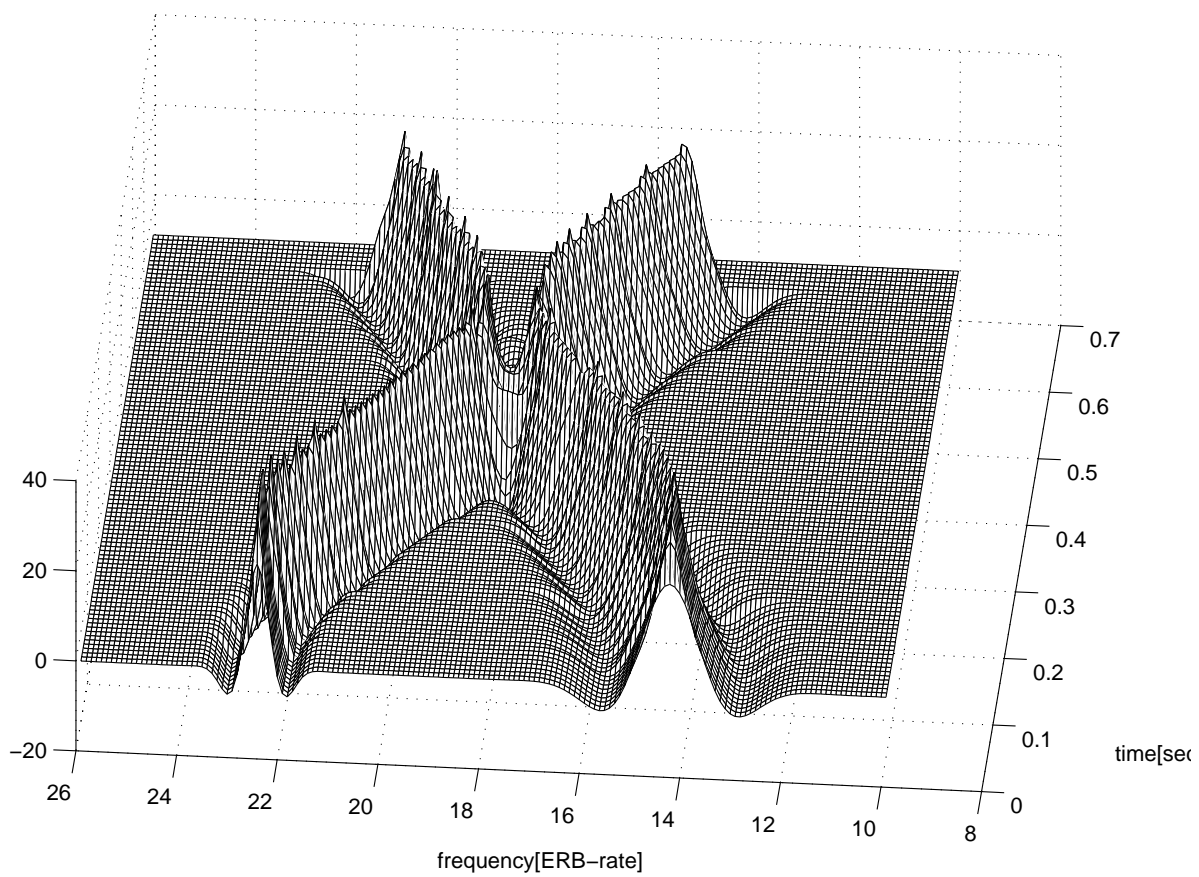


図 4.10: 交差するスイープ音の出力スペクトル



## 4.3 雑音による中断のある対数線形スイープ音に対するシミュレーション

### 4.3.1 目的

本モデルでは、雑音などによって目的信号がかき消されて存在しない場合でも過去の情報をもとに、予測追跡を行なうことができることを示す。(式(3.16)(3.20))また、図4.11の場合は滑らかに外挿したように知覚されるが、図4.12の場合は雑音内で周波数の変化方向が変わってしまうため、滑らかには知覚されない。このような違いも本モデルで模擬できることを示す。

### 4.3.2 シミュレーションに使用するデータ

Bregman の Demonstrations of Auditory Science Analysis[18] の中で紹介されている信号を用いる。これは、図4.11のように 1464 Hz ~ 2297 Hz へ 1000 msec で上昇するスイープ音に、500 msec を中心として 150 msec の区間を取り除く。そして、取り除いた部分にスイープ音の振幅レベルよりも大きい振幅レベルを持つ、長さ 150 msec の白色雑音を付加する。同様に、1000 ~ 2000 msec で 2297 Hz ~ 1464 Hz へ下降するスイープ音に 1500 msec を中心に同様の白色雑音を付加したスイープ音を作成した信号を 1 周期とするものを数秒間繰り返す。(今回は 1 周期 (2 sec) だけを用いた) この信号からケプストラム次数 60 次で不偏推定法より求めたスペクトル包絡を入力信号とした。

### 4.3.3 シミュレーション結果

図4.13の合成音を固有周波数を 4 Hz, 減衰定数を 1 として本モデルで予測・追跡した結果を図4.14に示す。雑音区間は、式(3.20)の時定数  $c = 0.0064$  とした。また、同条件で図4.12の合成音を予測・追跡したピーク周波数の結果を図4.15に示す。雑音内で変化方向が変わることによって、雑音外挿後、急に追跡方向を変えている様子を再現できている。

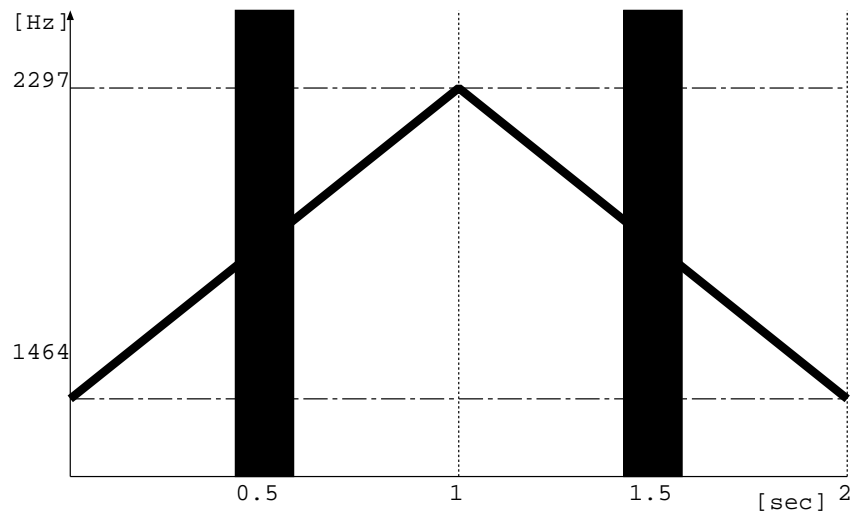


図 4.11: 雑音置換されたスイープ音 1

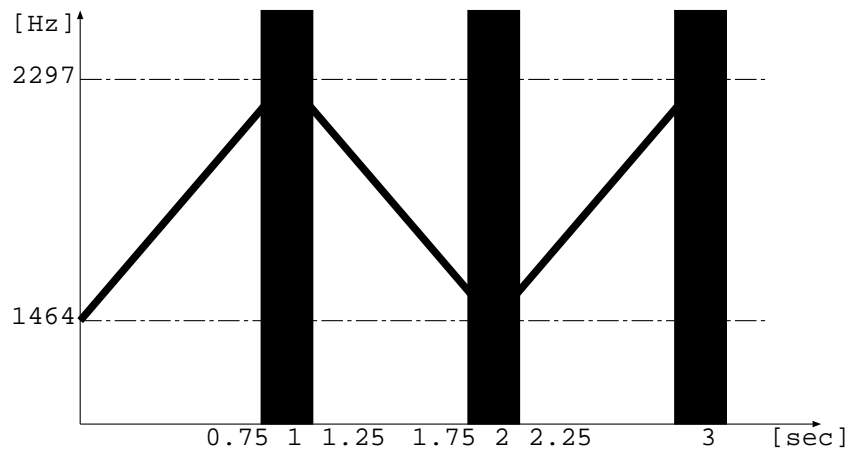


図 4.12: 雑音置換されたスイープ音 2

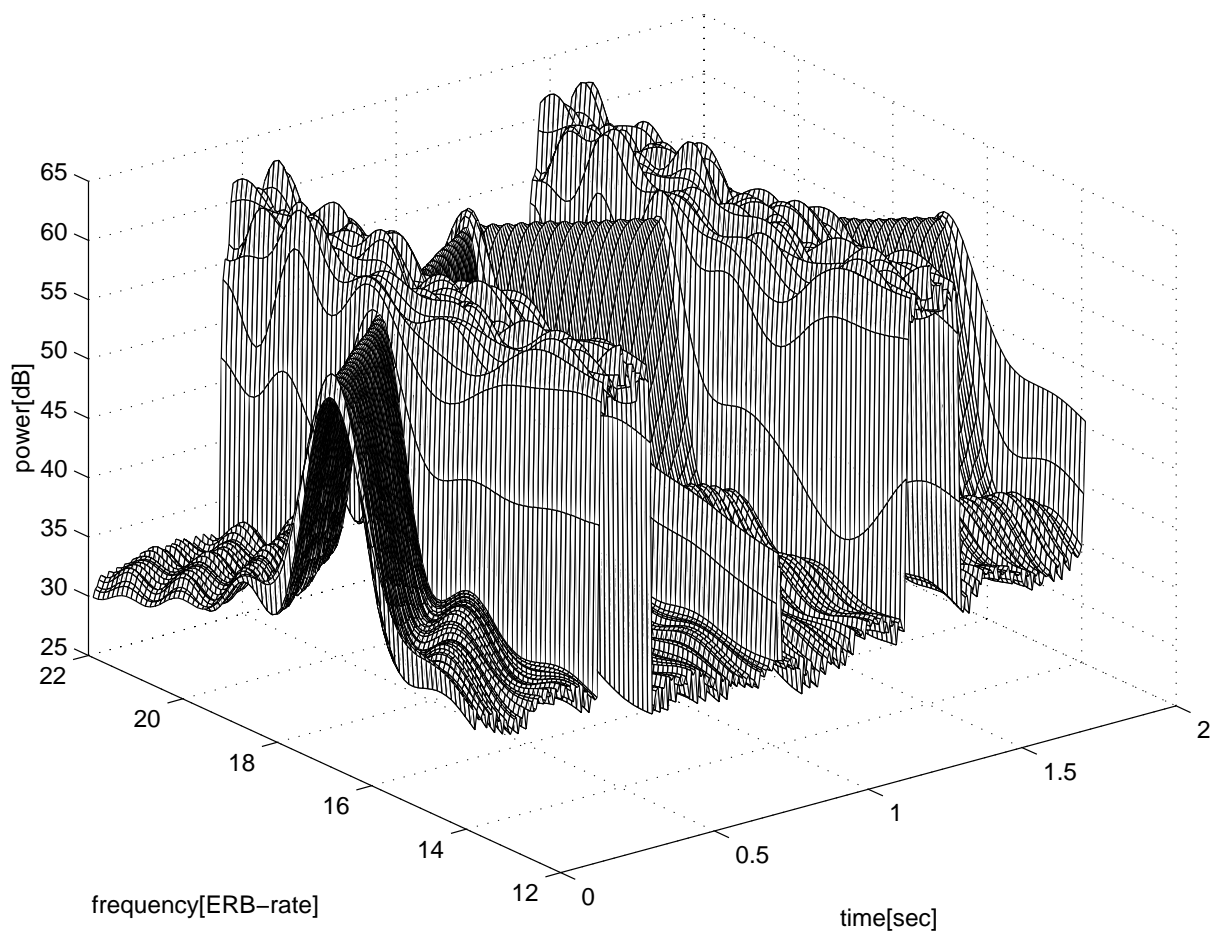


図 4.13: 雑音置換されたスイープ音の入カスペクトル

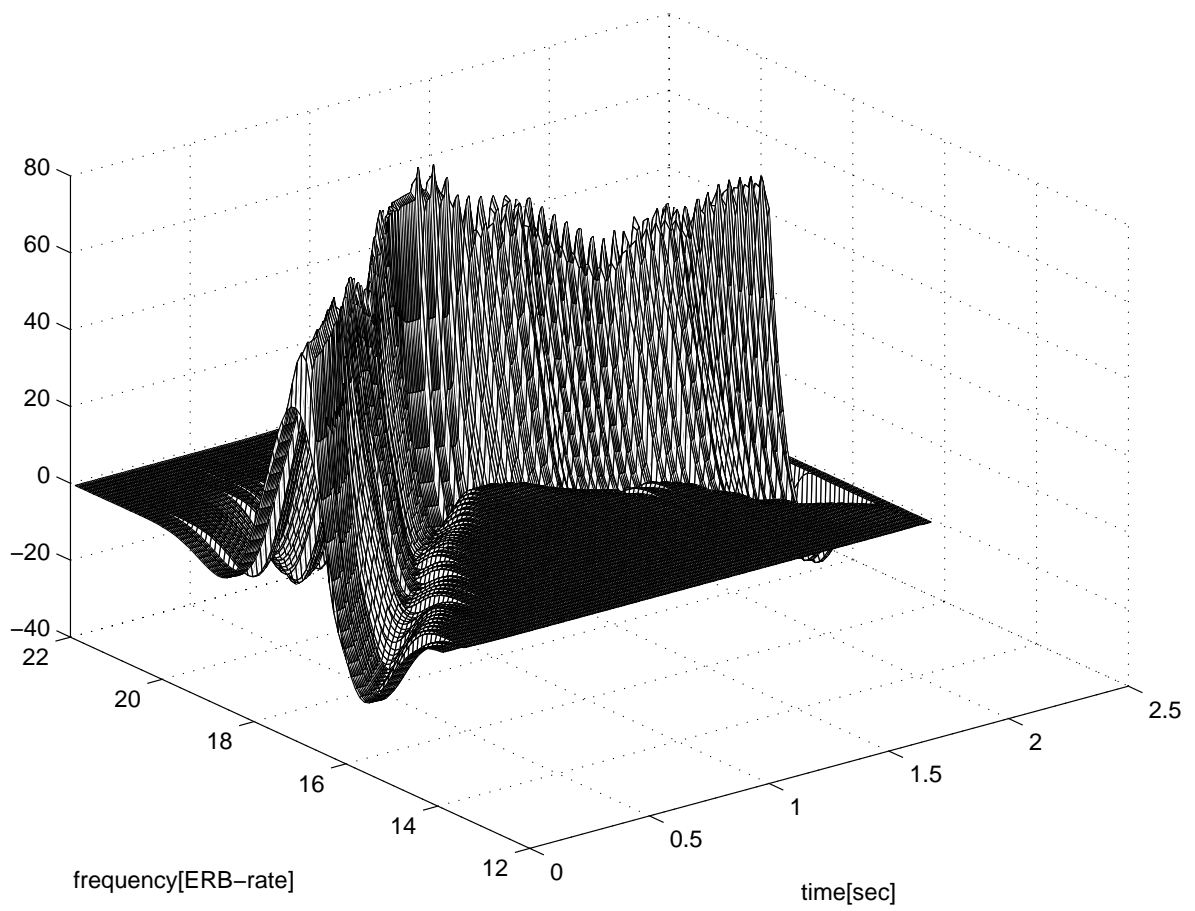


図 4.14: 雑音置換されたスイープ音の出力スペクトル

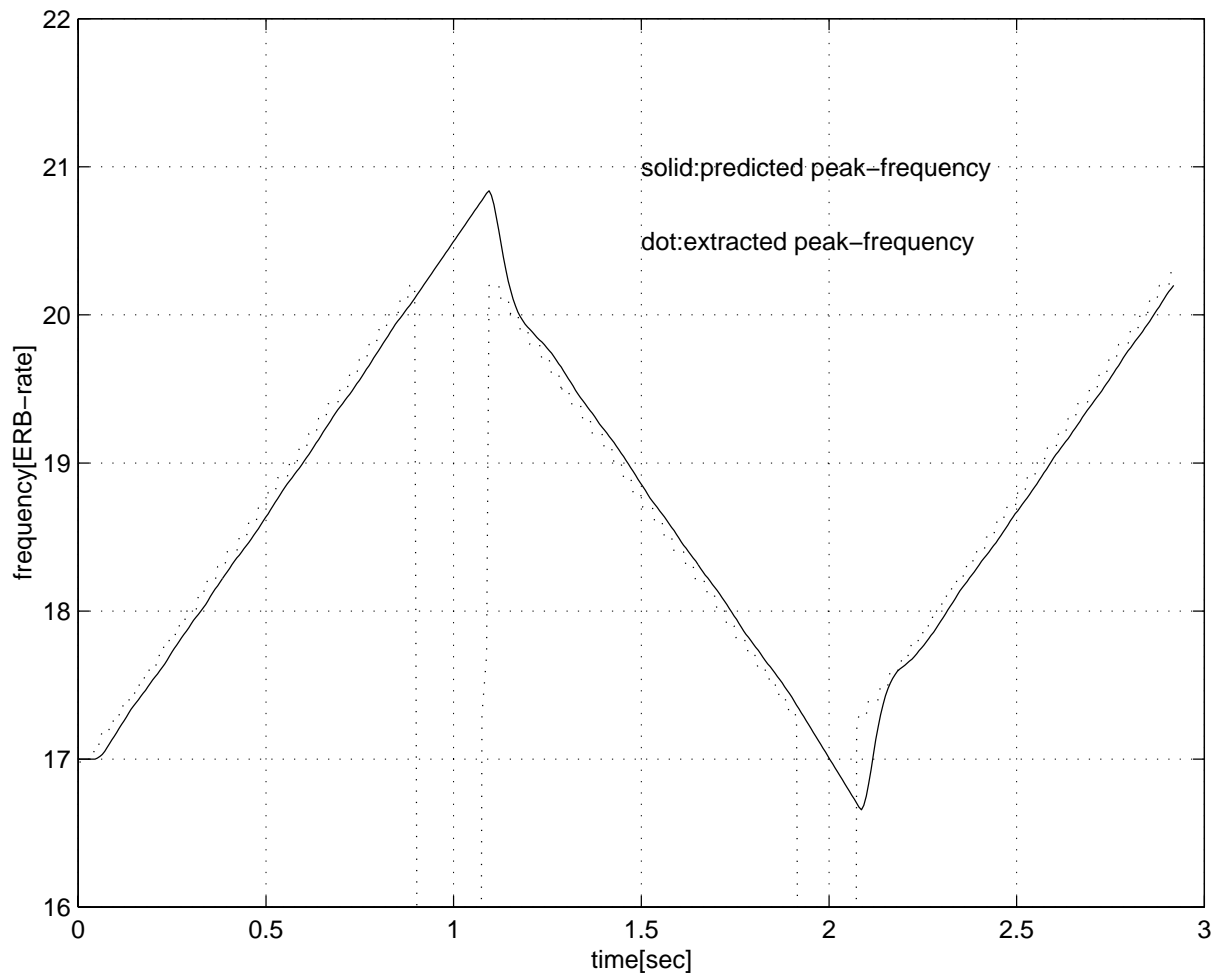


図 4.15: 雑音置換内で変化方向の変わる信号のピーク周波数の予測追跡 (実線: 予測追跡結果、点線: 抽出されたピーク情報)

#### 4.3.4 考察

本モデルが雑音内では雑音前の出力の傾き情報を利用し、予測追跡しているのに対して、河本や相川らの2次系システムによる予測追跡では、雑音内での予測に過去の出力情報のみを用いるため、雑音内を本モデルのように滑らかに外挿することができない。倉片らの報告からも、雑音内での知覚が雑音開始点と終了点を単純に結んだものではなく、過去の情報をバッファにため、未来の情報と過去の情報を融合させた、より複雑な知覚をしていると思われる。今回は過去の情報を用いた外挿のみであったが、今後内挿機能を加える必要があるだろう。

## 4.4 合成音声に対するシミュレーション

### 4.4.1 目的

連続音声中では声道の形が急には変化できないために、前後の音素の影響を受けて生じる「わたり音」の現象、また連続母音に対して、このホルマント周波数が前後の母音の影響を受けてなまける調音結合などが知られている。しかし、人間が知覚する場合、これらの不完全な音を補正し、知覚的に調音結合からの回復を行なっている。本モデルで、この現象を模擬するために合成母音を用いたシミュレーションを行なった。

### 4.4.2 シミュレーションに使用するデータ

萩原 [20] が行なった実験と同様の合成音を作成する。

ホルマントシンセサイザ [19] を使用して合成し、/a/の第1ホルマントは/u/と/a/の間ぐらいで音韻性を悪くしてある、連続合成母音/u/-/a/-/u/を用いた。合成音/a/はF1、F2 (622,1259 Hz)、/u/はF1、F2 (284,1259 Hz) を持ち、それぞれの帯域幅は80,120 Hzである。サンプリング周波数は20 kHz、基本周波数は140 Hz一定とした。第2ホルマントは固定し、第1ホルマントを遷移させる。各々の母音は120 msecであり、各々の遷移部は30 msecとした。(図4.16)この信号をケプストラム次数60次で不偏推定法より求めたスペクトル包絡を入力信号とした。

### 4.4.3 シミュレーション結果

図4.18の合成音を固有周波数を6 Hz、固有周波数0.0531として本モデルで予測・追跡した結果を図4.17、図4.19に示す。

### 4.4.4 考察

今回はシミュレーションデータとして萩原が行なった実験と同様のデータを用いた。萩原の実験から、このデータの場合0.6~0.8 ERBのovershoot量があると報告されている。図4.17から本モデルで予測追跡した結果、萩原により報告されている程度のovershootをし、なまけが知覚的に回復されていることを模擬できた。

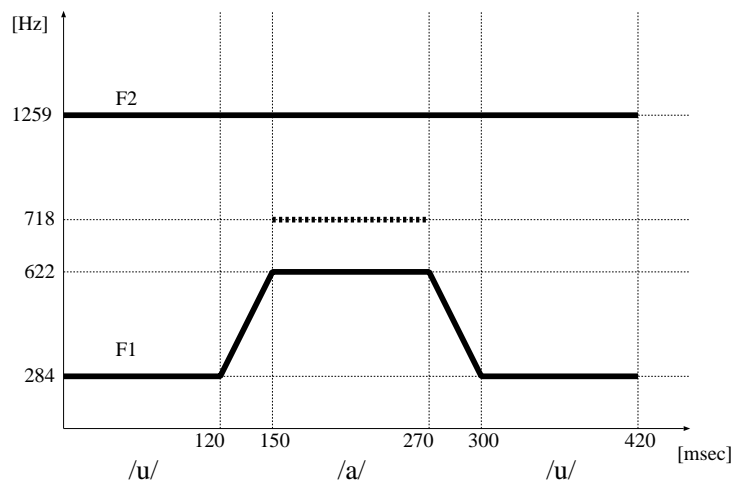


図 4.16: 連続合成母音 /u/-/a/-/u/

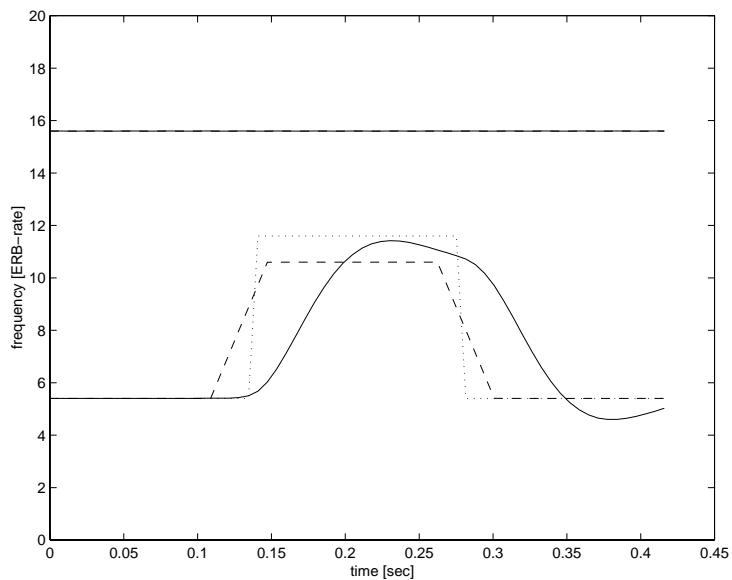


図 4.17: ピーク周波数の予測追跡結果 (実線: 予測追跡結果、破線: 原信号のピーク周波数、点線: 音韻性の良い/a/のピーク周波数)



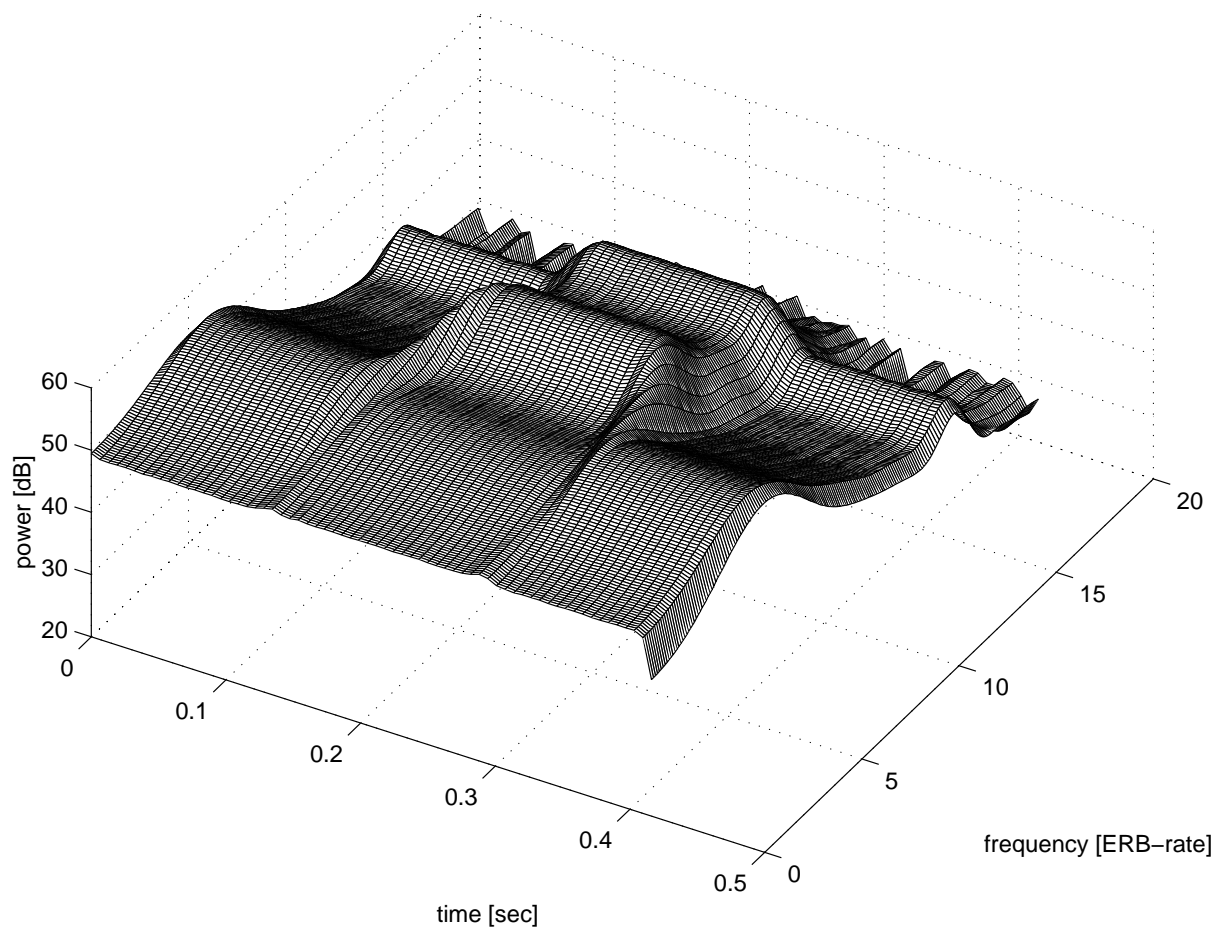


図 4.18: 連続合成母音の入力スペクトル

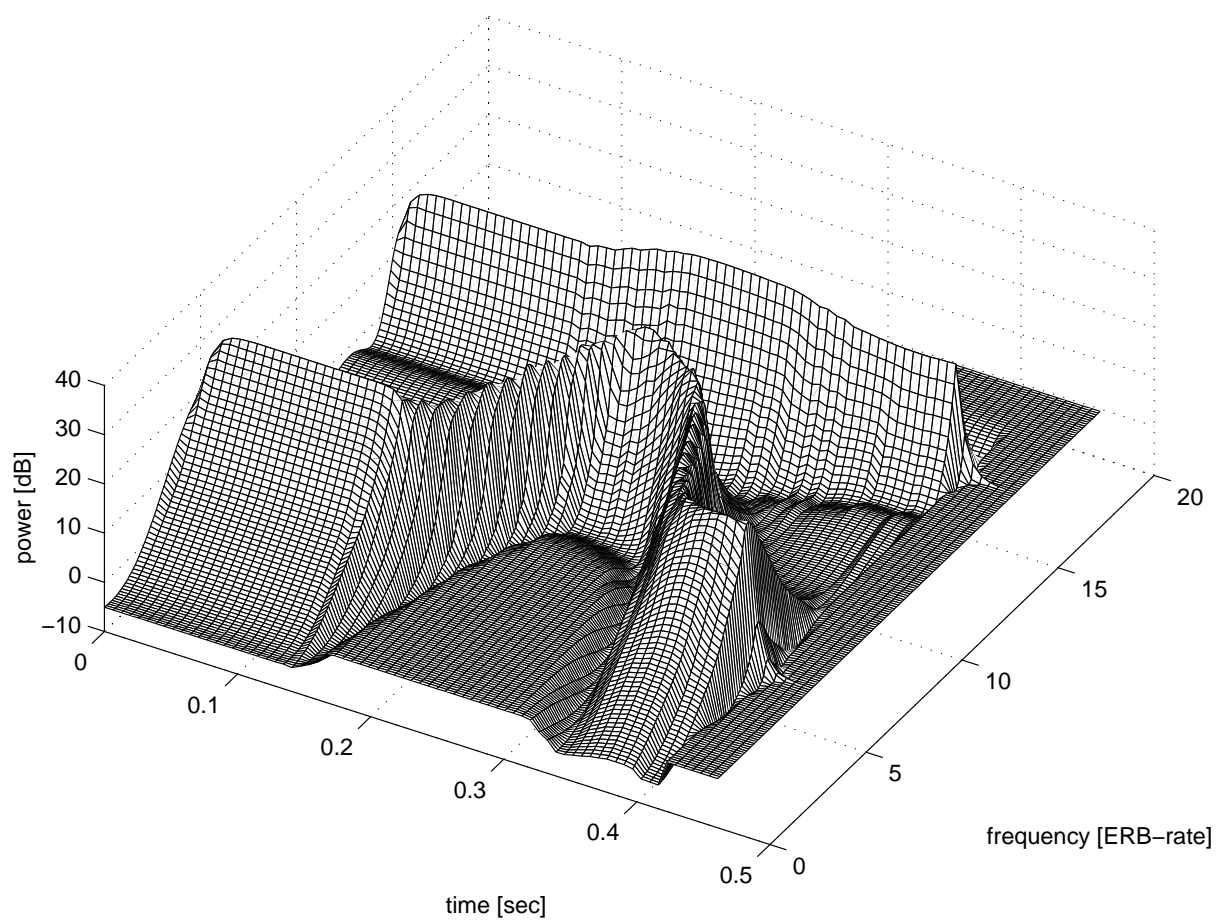


図 4.19: 連続合成母音の出力スペクトル

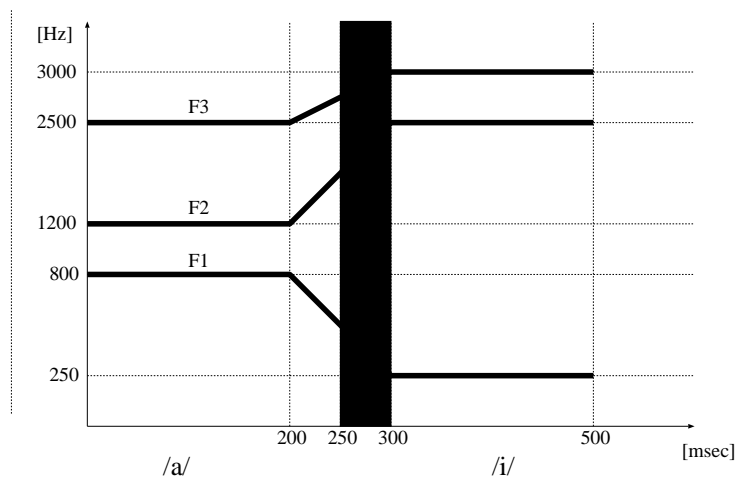


図 4.20: 遷移部が雑音置換された合成母音 /a/-/i/

## 4.5 遷移部が雑音置換された合成音声に対するシミュレーション

### 4.5.1 目的

音韻修復現象とは雑音によってかき消された音声を修復し知覚する現象である。本モデルにより予測追跡を行なうことにより、遷移部が雑音置換された合成音が音韻修復されることを示す。

### 4.5.2 シミュレーションに使用するデータ

ホルマントシンセサイザを使用して合成した、連続母音 /a/-/i/を用いた。合成音/a/は F1 ~ F3 (800,1200,2500 Hz)、/i/は F1 ~ F3 (250,2500,3000 Hz) を持ち、それぞれの帯域幅は 80,120,150 Hz である。サンプリング周波数は 20 kHz、基本周波数は 140 Hz 一定とした。/a/-/i/の 200 ~ 300 msec の遷移部の中心から終端までを長さ 50 msec の白色雑音で置換した。(図 4.20) この信号からケプストラム次数 60 次で不偏推定法より求めたスペクトル包絡(図 4.22)を入力信号とした。

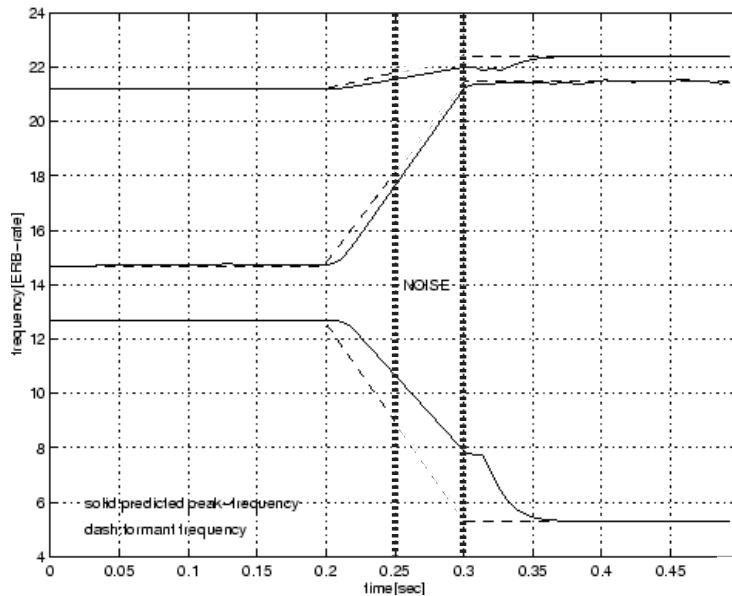


図 4.21: ピーク周波数の予測追跡結果（実線：予測追跡結果、点線：原信号のピーク周波数）

### 4.5.3 シミュレーション結果

図 4.22の合成音を固有周波数を 18 Hz, 減衰定数を 1 として本モデルで予測・追跡した結果を図 4.21、図 4.23に示す。本モデルを用いて予測追跡した結果、雑音によってかき消された信号を補完することによって、音韻修復現象を再現することが可能となった。

### 4.5.4 考察

本モデルを用いて予測追跡した結果、雑音によってかき消された信号を補完することによって、音韻修復現象を模擬することが可能となった。IFIS とは違って、スペクトルピークの局所的な情報のみを用いても音韻修復がシミュレートできた。今回のシミュレーションも雑音置換部のある線形対数スイープ音同様、未来の情報と過去を融合させた、内挿機能を加えより複雑な知覚を模擬していく必要があるだろう。

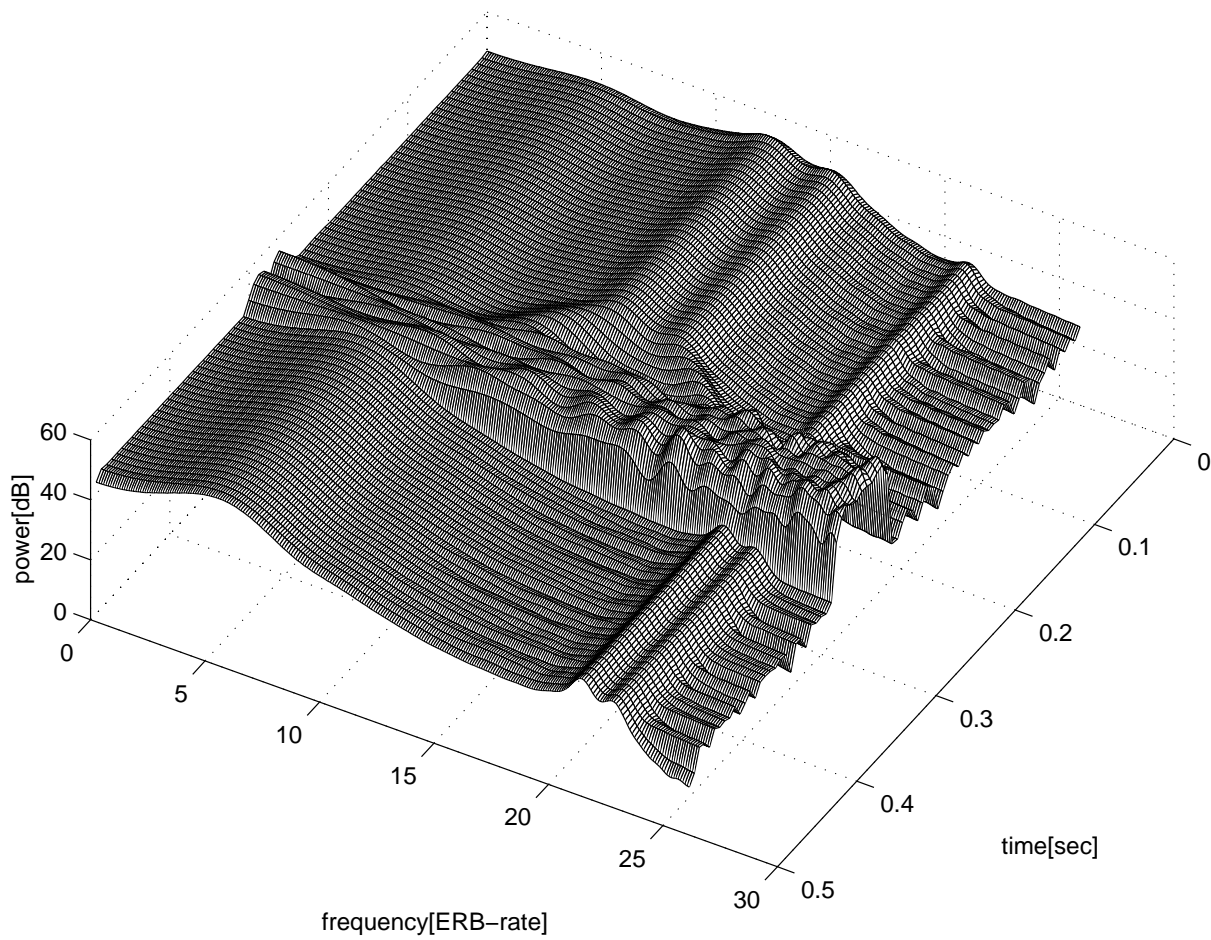


図 4.22: 遷移部が雑音置換された合成母音 (入力スペクトル)

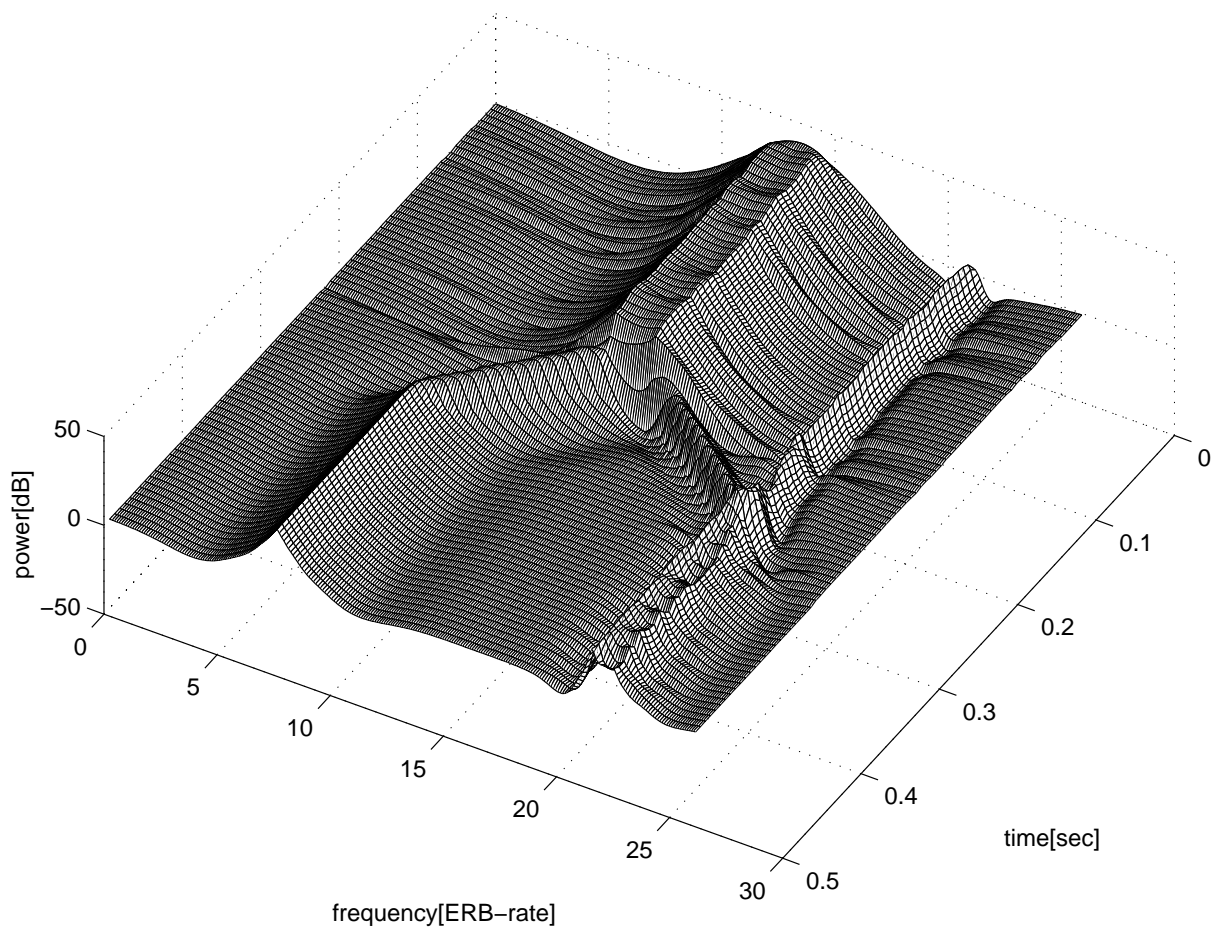


図 4.23: 遷移部が雑音置換された合成母音 (出力スペクトル)

# 第 5 章

## 結論

### 5.1 まとめ

本研究では、聴覚の音情報に対する補完現象をその一要因とされている周波数変化音に着目することによって、人間の聴覚の周波数変化音追跡機能を信号処理レベルにおいて実現できた。

Shamma らの 1 次聴覚野 (A1) モデルを用いてスペクトルピークを周波数、振幅、帯域幅、対称性で表現し、自己回帰直線を用いてピーク周波数を抽出した。抽出された各々のパラメータに対して新たに提案した 2 次系線形システムを用いて予測を行なった。この予測されたパラメータに Bregman の発見的規則の 1 つで、音脈の形成に関わる規則である「よい連続の原理や周波数近傍の変理の概念」を追跡系に導入し、システムの固有周波数、減衰定数を制御することによって複数のスペクトルピークの追跡が可能となり、種々の知覚現象を模擬することができた。相川、増田らの手法では不可能であった「スペクトル構造」を保持しつつ、周波数変化音を予測追跡することが可能となった。また、河本のモデルでは複数のピークを持つ信号を予測追跡すること、雑音置換された部分を判断しその置換部を予測追跡することなどが不可能であったが、今回のシミュレーション結果から、これらの問題点を解決することができた。

しかし、この固有周波数、減衰定数はシミュレーションを行ない、その過程から経験的に求めるに留まった。聴覚の予測追跡機能の存在を示唆する多くの研究が発表され結果が得られたが (例えば [20])、そのパラメータを定量的に決定するためには今後の研究を待たなくてはならないだろう。

本研究で行なった線形対数スweep音、合成音声に対するシミュレーション結果から、予測追跡システムのパラメータを変化させることによって、本モデルが信号処理レベルにおいて、聴覚の周波数変化音に対する知覚過程を近似することができるモデルであると検証できた。

## 5.2 今後の課題

人間の聴覚は、音韻を特徴付ける要因の一つである周波数変化を手がかりに音声の認識を行なっていると考えられる。今後、本モデルで実現した聴覚の周波数変化音予測追跡機能を音声情報処理へ応用するできれば、音声認識や音声合成などの高精度化、また聴覚情景解析の分野でも活用されるだろう。



# 謝辞

本研究を進めるにあたり、多大な御助言を頂き、熱心な御指導をして頂きました赤木正人助教授に心から感謝致します。

本研究に関して、多大な御助言、御討論をして頂いた岩城護助手に心より感謝致します。

本研究に関して、多大な御助言をして頂いた博士後期課程の鷓木 祐史氏、水町 光徳氏に心より感謝致します。

本研究を進めるにあたり、有意義な議論を共にして下さった赤木研究室の皆様方に心より感謝します。

## 参考文献

- [1] A. S. Bregman, Auditory Scene Analysis: The Perceptual Organization of Sound(MIT Press, Cambridge, MA. 1990).
- [2] Cooke, M. P. and Brown, G. J. 'Computational auditory scene analysis: Exploiting principles of perceived continuity,' Speech Communication 13 pp.391-399, 1993.
- [3] Brown, G. J. and Cooke, M. P. 'Computational auditory scene analysis,' Proceeding of ICSLP, pp.523-526, 1992.
- [4] Nakatani, T. , Kawabata, T. and Okuno, H. G. 'Unified Architecture for Auditory Scene Analysis and Spoken Language Processing,' Proceeding of ICSLP, pp.1403-1406, 1994.
- [5] D. P. W. Ellis, A Computer Implementation of Psychoacoustic Grouping Rules, proceeding of IEEE, pp.108-112, 1994.
- [6] 松井 理直, 倉片 憲治, 西村 篤. 予測処理課程を含む聴覚過程のモデル化, 日本音響学会講演論文集, I:443-444,1995.
- [7] 倉片 憲治, 松井 理直, 西村 篤. 周波数変化音に対する連続聴効果の知覚的軌跡, 聴覚研究会資料, H-94-71,1994.
- [8] 相川 清明, 津崎 実, 河原 英紀, スイープ音追跡系の動特性, 音響学会聴覚研資, H-95-31,1-8,1995.
- [9] 相川 清明, 河原 英紀, 周波数変化音追跡神経回路モデル, 信学技報, SP95-66,1-8,1995.
- [10] 増田 郁代, 相川 清明, 津崎 実, 周波数変化音追跡モデルに基づくスペクトル予測法, 信学技報, vol.96,SP96-2,pp.9-16,1996.

- [11] 河本 晃一, 周波数変化音に対する予測追跡の知覚モデルに関する研究, JAIST 修士論文, 1997.
- [12] K. Wang, S. A. Shamma, Spectral Shape Analysis in the Central Auditory System, IEEE Trans. Speech Audio Processing, vol.3, no.5, pp.382-395, 1995.
- [13] 今井 聖, 阿部 芳春, 改良ケプストラム法によるスペクトル包絡の抽出, 信学論, vol.J62-A, no.4, pp.217-223, 1979.
- [14] 今井 聖, 古市 千枝子, 対数スペクトルの不偏推定, 信学論, vol.J70-A, no.3, pp.471-480, 1987.
- [15] 赤木 正人, 聴覚フィルタとそのモデル, 信学会誌, vol.77, no.9, pp.948-956, 1994.
- [16] 佐藤 雅昭, ウェーブレット理論の数学的基礎 第 1 部 第 2 部, 音響学会誌, vol.47, no.6, pp.405-423, 1991.
- [17] 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.
- [18] A. S. Bregman, P. A. Ahad Demonstration of Auditory Scene Analysis : The Perceptual Organization of Sound(1995),
- [19] Klatt, D. H., Software for a cascade/parallel formant synthesizer, J. of Acoust. Society of America., vol.67, no.3, pp.971-995, 1980.
- [20] 萩原 力, 母音知覚における文脈効果モデルの高度化に関する研究, JAIST 修士論文, 1996.

## 学会発表リスト

- 坂口 博文、赤木正人、「スペクトルピーク追跡モデルを用いたスペクトル予測・追跡」、日本音響学会 平成 10 年度春季研究発表会 発表予定 ( 講演番号 1-8-14 )