

Title	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
Author(s)	河本, 晃一
Citation	
Issue Date	1997-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1017
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

修士論文

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

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

河本 晃一
情報研究科赤木研究室

1997年2月14日

目次

1	序論	1
1.1	背景	2
1.1.1	周波数変化の知覚	2
1.1.2	スペクトルの脳内表現	3
1.2	本研究の特色	3
2	スイープ音知覚と音声知覚	4
2.1	対数線形スイープ音の知覚に関する知見	4
2.2	周波数方向が変化するスイープ音知覚に関する知見	5
2.3	雑音中のスイープ音知覚に関する知見	6
2.4	スイープ音と音声の知覚	8
2.5	スイープ音の軌跡知覚に関する実験	9
3	1次聴覚野におけるスペクトル表現と機能モデル	11
3.1	1次聴覚野に対する入力	11
3.2	1次聴覚野における神経細胞群の応答特性	12
3.2.1	周波数応答特性	12
3.2.2	時間応答特性	13
3.3	スペクトル表現モデルにおける処理の特徴	13
4	周波数変化予測追跡モデル	15
4.1	入力スペクトル	17
4.2	1次聴覚野におけるスペクトル表現モデル	18
4.2.1	1次聴覚野細胞群における周波数応答の表現	18

a.	周波数応答特性の表現	18
b.	周波数に対する応答	19
4.2.2	1次聴覚野細胞群における時間応答の表現	21
a.	時間応答特性の表現	21
b.	時間的变化に対する応答	22
4.3	ピーク情報によるスペクトル表現	22
4.4	周波数変化に対する予測追跡機能	24
4.4.1	予測追跡機能の定式化	24
4.4.2	予測追跡機能の特性	26
4.4.3	雑音が付加された場合の予測	26
5	周波数変化予測追跡シミュレーション	28
5.1	対数線形スイープ音に対する予測追跡結果	28
5.2	段つき対数線形スイープ音に対するシミュレーション	30
5.3	雑音を後続させた対数線形スイープ音に対するシミュレーション	31
5.4	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	31
5.5	連続発話における母音のホルマント遷移を模倣した複数スイープ音に対する予測追跡結果	34
6	結論	37

図一覽

2.1	相川らによる段つき対数線形スweep音の軌跡知覚におけるピッチの揺らぎ [4]	5
2.2	倉片らによる雑音により中断された対数線形スweep音の軌跡知覚 [10]	7
2.3	母音発声時におけるホルマント軌跡の調音結合による影響 (概略図)	8
2.4	対数線形スweep音の軌跡知覚実験における解答例	10
4.1	周波数変化予測追跡モデルのフローチャート	16
4.2	入力スペクトル例	17
4.3	応答関数の概形	18
4.4	入力スペクトル包絡とその静的スペクトル表現	20
4.5	時間-周波数応答関数の概形	21
4.6	スペクトル再構成結果	23
4.7	2次予測追跡機能の特性	27
5.1	対数線形スweep音に対する予測追跡	29
5.2	段つき対数線形スweep音に対する予測追跡	30
5.3	雑音を後続させた対数線形スweep音に対する予測追跡	32
5.4	雑音による中断のある対数線形スweep音に対する予測追跡 1	33
5.5	雑音による中断のある対数線形スweep音に対する予測追跡 2	35
5.6	連続母音のホルマント遷移を模倣した複数スweep音に対する予測追跡	36

第 1 章

序論

音情報が雑音などによって欠落するような場合、計算機による音声認識は難しい。しかしこのような環境下においても、人間は音韻修復やカクテルパーティ効果などのように欠落した情報を補完し音声によるコミュニケーションを正確に行なうことができる。このような補完現象は、知覚対象とする音を連続した1つの音脈として知覚し、その音脈において欠落している情報を補うことにより生起するものと考えられる。また、音脈の形成において時間的周波数変化の連続性がその要因の1つあり、このことから聴覚には周波数変化を予測し追跡する機能の存在が示唆される。

このような音情報に対する人間の補完機能を信号処理レベルにおいて実現することにより、音声認識の高度化を図ることが可能と考えられる。音響信号は、聴覚末梢において周波数分析されるため、その周波数が変化するのにもない音響スペクトルが変化する。人間はスペクトルにおける変化情報を用いて一連の音脈を形成しているものと考えられる。そこで本研究においては、周波数変化にともなうスペクトルの変化に着目し、周波数変化音を表現するのに適したスペクトル表現の検討、およびそのスペクトル表現を用いた周波数変化音に対する予測追跡機能の信号処理レベルにおける実現を試みた。

本論文において、以下第1章において背景と本研究の特色、第2章において対数線形スイープ音の知覚現象に関する実験及び他の研究からの知見、さらにスイープ音のピークと音声知覚との関係について述べる。第3章においては1次聴覚野におけるスペクトル表現モデルの説明、また第4章では周波数変化予測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。さらに第5章において単一ピークからなる様々な周波数変化音、および連続母音発話時のホルマント遷移を模倣した複数の周波数変化音を対象としたシミュ

レーション結果を述べ考察する。

1.1 背景

1.1.1 周波数変化の知覚

1990年にBregmanによって聴覚の初期段階における処理に関する心理物理学的知見が整理された[1]。それを機に、一連の音刺激を1つのまとまった音脈として捉える streaming あるいは grouping の効果とその形成機構に研究の焦点があてられるようになり、心理学的側面だけでなく、生理学的及び工学的側面からの聴覚における音環境理解の過程に関する研究が盛んになった。

周波数変化の知覚に焦点を当てた研究として、相川ら[4]は、時間に対して周波数が対数的に変化する対数線形スイープ音の知覚が、その始端部では変化が緩やかであり、次第に変化が速くなる湾曲した形状に知覚されることを報告している。さらに、途中で周波数停留部を含む段つきスイープ音に対する知覚を対象として心理物理実験を行なった結果、周波数変化の知覚現象が2次の伝達特性を持つシステムによってモデル化できることを報告している。

また、増田ら[6]はスペクトルが変化する合成母音に雑音が後続する刺激音の音韻修復に関する聴取実験を行なった。その結果、音韻修復現象には言語的な知識だけでなく、音の軌跡を追跡する非言語的な過程の音響処理による予測が行なわれていると結論付けている。

さらに周波数変化追跡機能の工学的応用について、増田ら[5]が相川らの周波数変化音追跡モデルを拡張し、音声のスペクトルパターンを対象として、雑音置換された部分に知覚されるスペクトルパターンを予測する手法を提案している。しかしながら、これらのモデルにおいては物理パラメータや手法などの生理学的及び心理学的な知見に対する妥当性については考察されていない。

倉片ら[10]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。倉片らは途中で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

数を結んで単純に補間するのではなく、さらに複雑な連続知覚が行なわれており、その1要因として周波数変化を追跡、予測する聴覚機能が存在する可能性を報告している。

1.1.2 スペクトルの脳内表現

Shamma ら [19] は1次聴覚野における応答領域の構造及びその機能に関する生理学的な実験結果を報告している。それによると、1次聴覚野には等周波数平面に沿って周波数変化に反応する応答領域が存在し、それぞれの応答領域は任意の周波数におけるスペクトルの局所的な傾きを符合化する機能を有している。その結果、1次聴覚野における空間位相選択性と同様に、1次聴覚野はスペクトル包絡においてピーク周辺を強調する働きを持っていることがわかった。

この1次聴覚野に対する生理学的な知見をもとに、Wang と Shamma[21] は1次聴覚野におけるスペクトルを周波数的な位置、帯域幅、対称性の3つの特徴的な軸によって表す知覚モデルを提案している。このモデルは1次聴覚野において音の基礎分析に用いられる振幅と位相の情報を保持しており、生理学的知見を考慮したスペクトル表現モデルであると言える。

1.2 本研究の特色

本研究では、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、Wang と Shamma の1次聴覚野におけるスペクトル表現の知覚モデルを用いてスペクトルを表現し、そのスペクトルに対して外挿及び内挿する機能を付加することによって周波数変化音の予測追跡機能をモデル化する。ここで、1次聴覚野におけるスペクトル表現を用いることにより、スペクトルピークを振幅、周波数、帯域幅、位相の4パラメータにより表現でき、さらにスペクトルの概形を各ピークの情報のみで表現することが可能となる。このような点から、スペクトルにおける情報を圧縮して表現していると言える。

また構築した予測追跡モデルを用いてシミュレーションを行ない、対数線形スイープ音の湾曲知覚現象や音韻修復現象などの心理物理実験の結果より得られている知覚現象を信号処理レベルにおいて工学的に実現する。

第 2 章

スweep音知覚と音声知覚

2.1 対数線形スweep音の知覚に関する知見

スweep音 (FM 音) の知覚像については、それを観測する方法論を含め様々な議論が存在する。

最近の研究では、相川ら [2] によりスweep音追跡系動作に関する知見が報告されている。1kHz から 3kHz を 600msec で変化する対数線形スweep音を刺激音とし、14 名の被験者に対して心理物理実験を行った結果、周波数が対数的に直線に上昇もしくは下降する対数線形スweep音の周波数変化が湾曲して知覚されることが確認された。

さらに周波数変化を円弧状に湾曲させたスweep音を用いて同様の実験を行ない、直線知覚に関する考察を行なっている。1kHz から 3kHz を 600msec で変化する対数線形スweep音を基準 (0%) に終端周波数側を正として、始端周波数と終端周波数の中間をスweep音が横切る時間が-50%、-25%、0%、25%、50%となるスweep音を 15 名の被験者に呈示し、周波数がより直線的に知覚できる刺激音を回答させた。その結果、-25% の場合、すなわち対数線形スweep音知覚における湾曲とは逆側に湾曲させたスweep音が直線知覚されており、このことから、対数線形スweep音が湾曲して知覚されていることが裏付けられる。

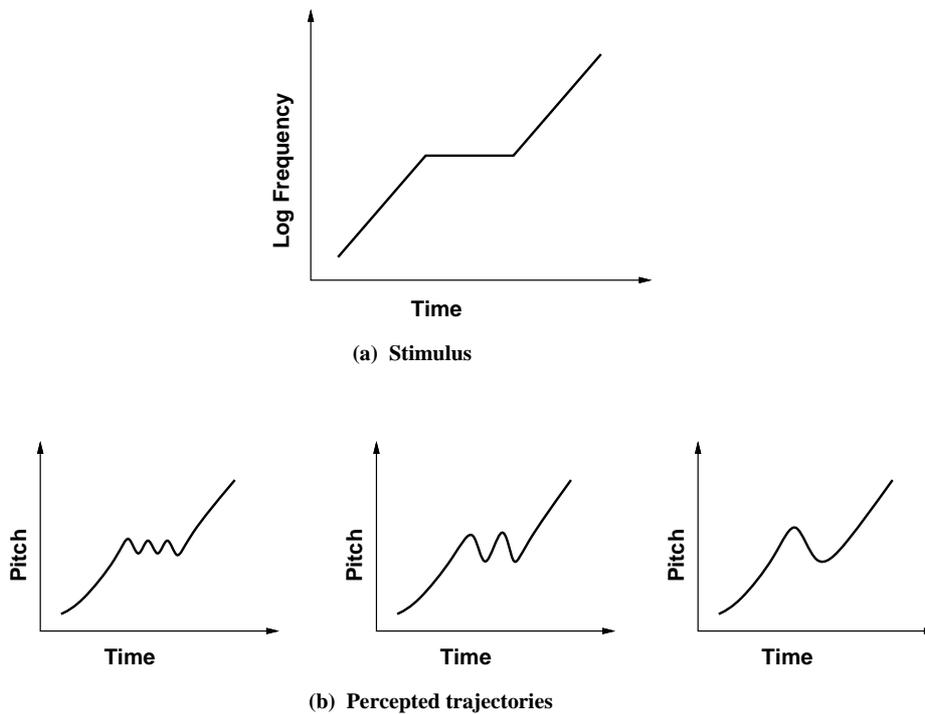


Figure 2.1: 相川らによる段つき対数線形スイープ音の軌跡知覚におけるピッチの揺らぎ [4]

2.2 周波数方向が変化するスイープ音知覚に関する知見

また相川ら [3][4] は、途中に周波数変化のない停留部を含む段つきの対数線形スイープ音を被験者に呈示し、知覚した軌跡を回答させる実験を行なっている。呈示したスイープ音の概形及び回答を整理したものを Figure 2.1 に示す。

この実験の結果、周波数停留部において Figure 2.1 のようにピッチの揺らぎが生じることがわかった。相川らはこの知覚軌跡を 2 次系のシステムを用いて再現可能であることを示した。さらにその逆フィルタを構成しそれを段つきスイープ音の周波数軌跡に対して作用させた刺激音を用いて心理物理実験を行なった結果、その知覚においてピッチの揺らぎが抑制されることを報告している。以上のことは、周波数追跡の特性は 2 次系システムによって近似できることを示唆している。

2.3 雑音中のスイープ音知覚に関する知見

また倉片ら [10] は、周波数変化音に対する連続聴効果について心理物理実験を行ない、雑音中における音に対する知覚的軌跡は雑音の開始点と終了点の周波数を結んで単純に補間するのではなく、複雑な連続性の知覚が行なわれていることを示す結果が報告している。

倉片らは、1k から 2kHz の周波数帯域を 500msec で上昇及び下降する対数線形スイープ音の直後に雑音を後続させた場合と、雑音を後続させなかった場合とのスイープ音の主観的な持続時間の測定を行なった。被験者は 8 名であり、後続させる雑音は 100、200、400msec の持続時間のもを用いた。その結果、雑音を後続させた場合は、雑音を後続させなかった場合に比べスイープ音の持続時間を過大評価すること、さらに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ことがわかった。これはスイープ音が後続雑音の中に延長されているかのように知覚されていることを示唆している。

また、途中で雑音が付加された対数線形スイープ音に対する知覚について実験を行なった。2kHz から 1kHz に 500msec で対数線形的に下降するスイープ音¹に 300msec の雑音を後続させ、その後に始端周波数と終端周波数の異なる各種の上昇及び下降スイープ音をさらに後続させた。それらの刺激音を 20 名の被験者に呈示し、雑音及び後続するスイープ音に対して知覚した軌跡を線画で表現させる作図法を用いて回答させた。倉片らが用いた刺激音の概形、およびそれらに対する被験者の回答のそれぞれ主なものを Figure 2.2 に示す。

この実験において、被験者には雑音中のスイープ音の有無は知らされていない。それにも関わらず、Figure 2.2のように、スイープ音の途中で雑音中にスイープ音が存在するかのように回答している。

また、Figure 2.2における A の場合、雑音終了部もしくは雑音中央部において折れ曲がる回答が全体の 90% を占める。また、B においても雑音終了部もしくは雑音中央部までスイープ音を外挿して知覚するパターンが 65% であり、C においても雑音以前のスイープ音をまっすぐに外挿する知覚パターンと雑音中において軌跡が S 字型に曲がる知覚パターンが 90% を占める。これらの実験結果は、周波数変化音の直後に雑音が付加された場合に雑音中を外挿して知覚する傾向があることを示している。

¹倉片らは 500Hz から 1kHz に 500msec で対数線形的に上昇する音を用いた同様の実験を行なっている。その結果も下降音を用いた場合とほぼ同等な傾向が確認されている。

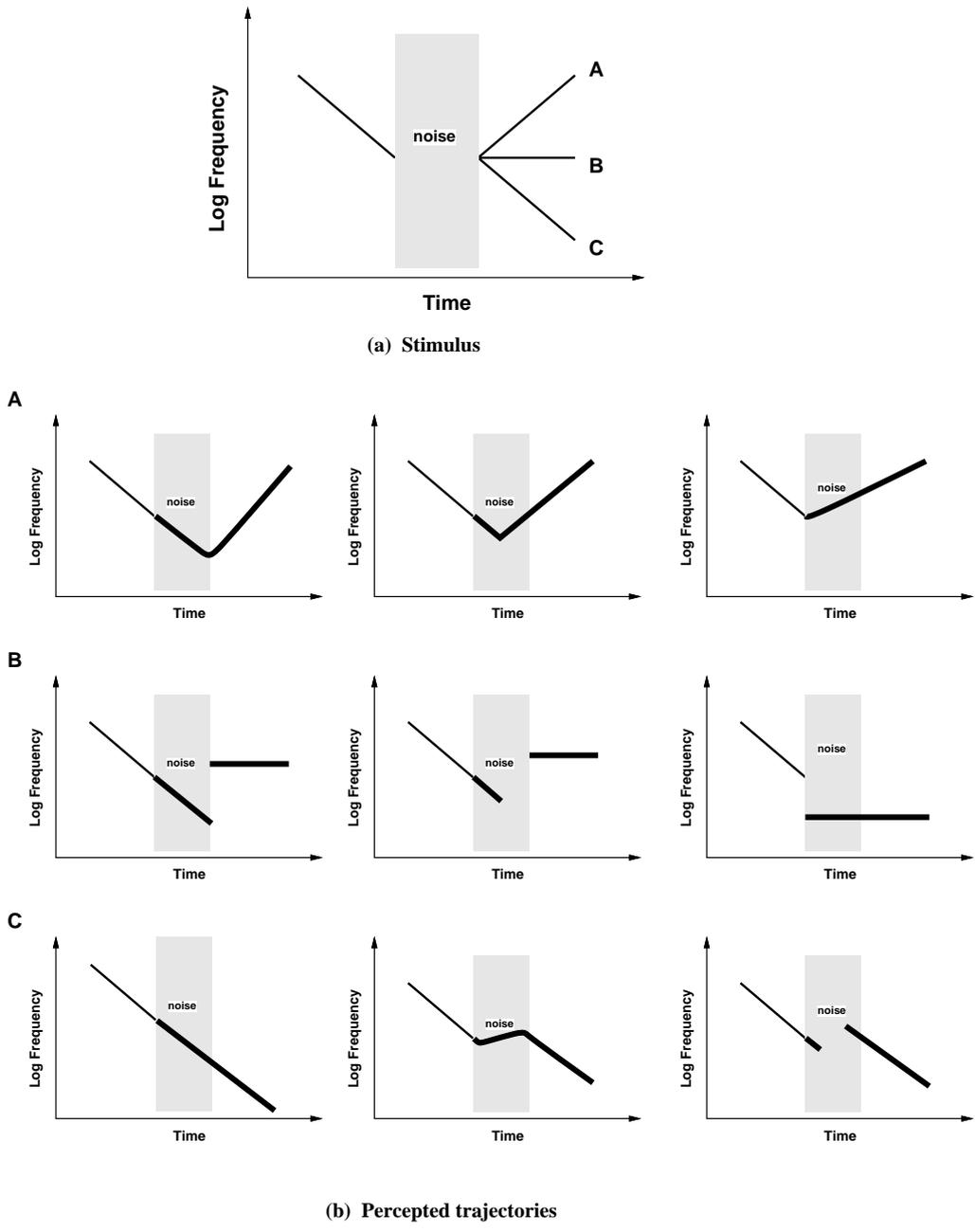


Figure 2.2: 倉片らによる雑音により中断された対数線形スイープ音の軌跡知覚 [10]

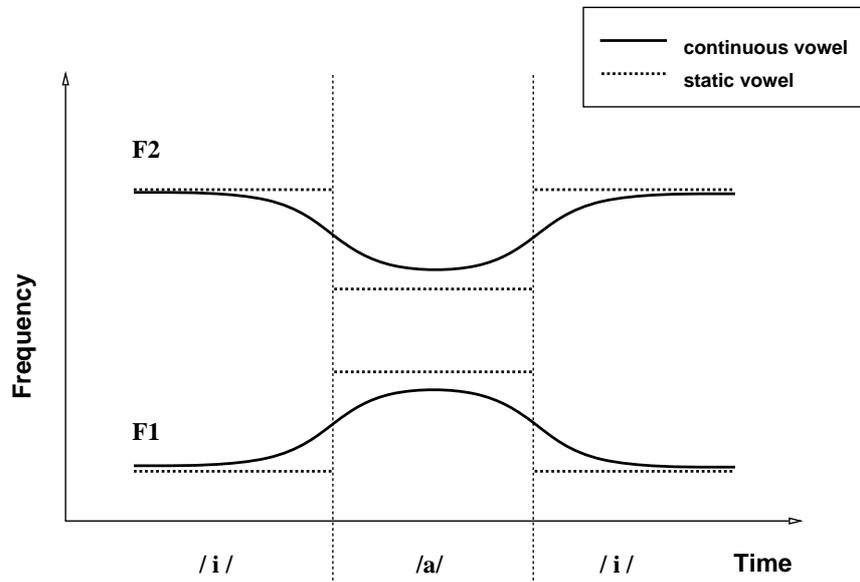


Figure 2.3: 母音発声時におけるホルマント軌跡の調音結合による影響 (概略図)

単独発声 (/i/ /a/ /i/) と連続発声 (/iai/) における第 1 および第 2 ホルマントの軌跡

倉片らは、これらの実験から、複雑な連続性知覚を生起する要因の 1 つとして周波数変化を追跡しその後の軌跡を予測する聴覚機能が存在する可能性を報告している。また松井ら [11] は、倉片らの心理物理実験の結果を用いて聴覚的情景分析における心理的過程のモデル化を試みている。

2.4 スイープ音と音声の知覚

人間が音声を知覚する場合には、スペクトルの局所的なピークの時間的遷移に関する情報を用いていると考えられる。特に単母音の知覚においては局所的なピーク、すなわち各音韻を特徴付ける共振周波数であるホルマントの分布の相対的位置関係が音韻の認識に用いられている。しかし連続発話においては調音器官の制約により各音韻が独立して生成されないためにホルマント周波数がずれる調音結合という現象が生じる [7]。例として「いあい」の単独発話及び連続発話時のホルマント軌跡の概略を Figure 2.3 に示す。調音結合によって生じる音響的特徴が不完全な音を人間が知覚する場合、動的音響特徴を追跡しその情報を用いて不完全な音響特性を補正するものと考えられる。

また、子音の知覚においては、特に破裂音の弁別においてその前後に続く母音との間に生じるホルマント遷移の方向と速度が手がかりとなっている。van Wieringen と Pols[12] は、単一及び複数のホルマントが存在する疑似ホルマント音を用いて、子音-母音 (CV) と母音-子音 (VC) における遷移部の知覚弁別について心理物理実験を行ない、遷移の変化率と弁別可能な周波数帯域の関係に関して考察している。

/b/と/d/における第2ホルマントの遷移は、その始端周波数と後続もしくは先行する母音のホルマント周波数の違いに生じ、人間はこのホルマント遷移を手がかりとしてこれら2つの子音を弁別している。van Wieringen と Pols は/b/ と/d/の始端周波数をそれぞれ700Hz、1800Hzとし、それらの子音と2種類の母音/u/と/a/ (第2ホルマント周波数はそれぞれ800Hz、1300Hz)を組み合わせ、/ba/、/bu/、/ab/、/ub/の疑似ホルマントを合成器により作成した。これを基準の刺激音とし、子音部におけるホルマント遷移の持続時間や周波数範囲を変化させ、基準の刺激音との弁別閾について実験を行なった。

その結果、遷移部の持続時間が長くなることにより遷移は弁別しやすくなり、遷移する周波数帯域が広いほど弁別は困難となる。また母音の位置が子音よりも先行するVCよりも後続するCVのほうが弁別しやすい。さらに、遷移部の上昇及び下降の違いや母音のホルマント周波数による違い、遷移の変化を基準よりも急にするか緩やかにするかの違いについては、弁別閾の変化に対する要因とはならないことが報告されている。これらの結果を踏まえ、van Wieringen と Pols は、周波数変化の弁別において、遷移する周波数帯域の広さや持続時間、それにとともなう変化率などが弁別閾の変化に対し関係があると結論付けている。

このように、音韻を特徴付ける要因の1つとして周波数変化があり、人間は音響信号におけるこれらの周波数変化を捉え、それを手がかりに音声の認識を行なっている。このような意味において、周波数方向もしくは時間方向の局所的な周波数変化は音声認識における重要な手がかりであり、計算機による音声認識に対して本研究の周波数変化音予測追跡モデルを用いることにより高精度化が可能と考えられる。

2.5 スイープ音の軌跡知覚に関する実験

本研究においても、スイープ音知覚における軌跡の形状についての予備的な心理物理実験を行ない、その知覚における軌跡の形状について検討を行なった。

本実験に用いた刺激音は、1kHz から 3kHz までの周波数帯域を 600msec で上昇する対

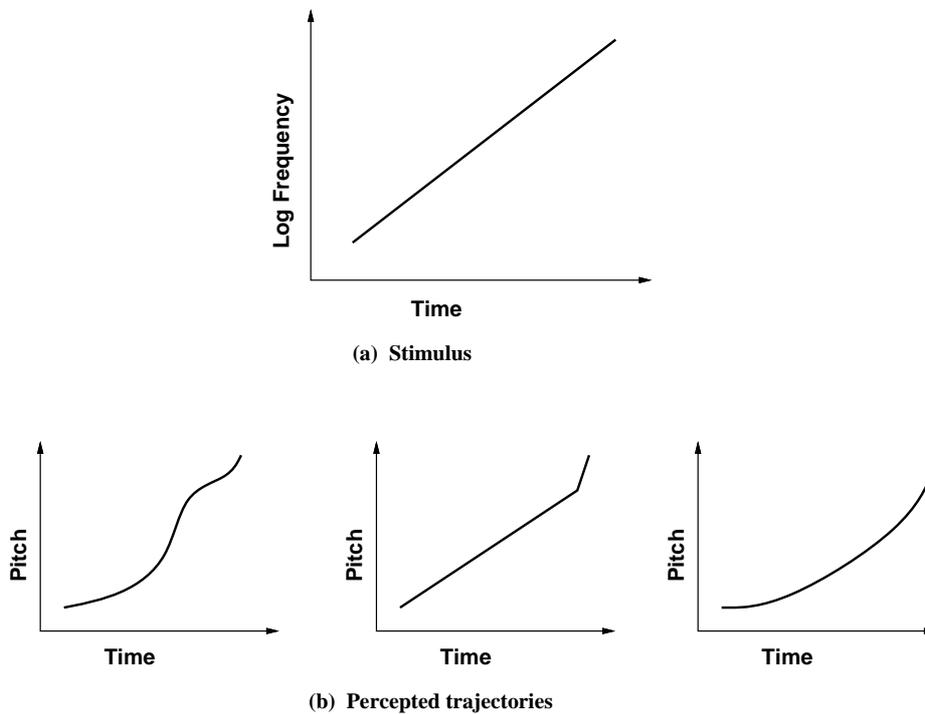


Figure 2.4: 対数線形スイープ音の軌跡知覚実験における解答例

刺激音：1kHz から 3kHz までを 600msec で対数変化する上昇スイープ音

数線形スイープ音である。女性 1 名を含む 5 名の被験者に対して、ヘッドホンを用いて刺激音を呈示し、その軌跡の概形を作図させ回答させた。また刺激音の呈示においては被験者による聞き直しを許した。実験に用いた刺激音の概形とそれに対する被験者の回答を整理したものを Figure 2.4 に示す。

軌跡知覚における概形や始端音及び終端音の高さを定量的に評価することは、結果の分散が大きかつ被験者の回答が普遍ではないことから困難であった。しかし、Figure 2.4 に示すようにほとんどの被験者の回答はスイープ音の変化が遅れる方向に湾曲して知覚している傾向を示していることが確認できる。この結果は相川ら [3] による周波数変化音の動的知覚特性に関する実験結果を支持するものである。

また、スイープ音の持続時間を 1200msec とした場合についても実験を行なったが、600msec と同様な結果が得られた。

第 3 章

1 次聴覚野におけるスペクトル表現と機能モデル

本研究で構築する周波数変化音予測追跡モデルにおいては、音響スペクトルの表現手法として Wang と Shamma[21] らによって提案された 1 次聴覚野におけるスペクトル表現モデルを用いる。

このスペクトル表現モデルは、Shamma ら [19] による 1 次聴覚野応答領域の構造及びその機能に関する生理学的調査実験の結果をもとにしている。以下の節において、Wang と Shamma らの 1 次聴覚野におけるスペクトル表現モデルを生理学的知見とあわせて説明する。

3.1 1 次聴覚野に対する入力

聴覚に入力された音響信号は、聴覚末梢系において雑音などによる変形に対して頑健でありかつ強調されたスペクトルに変換される。Wang と Shamma[20] は、基底膜における周波数分析、有毛細胞における情報変換、聴神経における側抑制による変形の 3 つの処理をモデル化し、入力音響信号に対する蝸牛からの出力をこれらのモデルから求めることができることを示している。この聴覚初期の機能モデルは、フレーム処理に基づいた短時間パワースペクトルを自己正規化することにより、スペクトルの雑音による変形の抑制、またスペクトルピークの強調などの特徴を持つ。

1 次聴覚野スペクトル表現モデルでは、上オリーブ複合体や下丘などは信号中継のみを

行なう中間機構として考えており、入力として前述の蝸牛から聴神経までを模擬した聴覚初期機能モデルを用いて求めた音響スペクトルを採用している。

3.2 1 次聴覚野における神経細胞群の応答特性

視覚や体性感覚と同様に、1 次聴覚野の神経細胞は刺激の種類に対し応答選択性を有し、各細胞はある特定の周波数、帯域幅、対称性を持つスペクトルピークに対して強く反応する。このことから、1 次聴覚野においてスペクトルピークの周波数的な位置と鋭さ、ピーク周辺のスペクトル形状を解析し、また表現しているものと解釈することができる。

Shamma ら [19] によると、1 次聴覚野においては特徴周波数の等しい細胞がまとまって構成される傾向がある。それらを 1 つの細胞群として考えると、それぞれの細胞群は特徴周波数付近に局在した応答特性を有し、その応答特性の対称性と帯域幅によって異なった機能を持つことがわかっている。これらの細胞群は、周波数、帯域幅、対称性のそれぞれ直交した 3 つの軸から成るコラム構造を形成している。

これらのことより、1 次聴覚野におけるスペクトル表現には多次元座標軸機構を採用していると見なすことができる。すなわち、任意の時刻における神経細胞群の応答 r は、周波数 b 、帯域幅 a 、対称性 (位相) ϕ の 3 パラメータにより表現することができる。

3.2.1 周波数応答特性

各神経細胞群は、特定の帯域幅と対称性を持ち、ある周波数に局在した周波数応答特性を持つ。神経細胞群の周波数応答特性を近似する周波数応答関数 ψ は、次式のように、周波数 b 、帯域幅 a 、対称性 ϕ の 3 つのパラメータによって表現される。

$$\psi = f(a, b, \phi)$$

Wang と Shamma は神経細胞群の周波数方向への側抑制を考慮し、応答関数 ψ として Laplacian-Gaussian 関数 $\nabla^2 G(b, a, \phi)$ を採用している。その帯域幅 a をスケーリングパラメータにより制御し、また対称性 ϕ についてはヒルベルト変換により定義された対称関数と非対称関数の和として表される。

1 次聴覚野における細胞群の周波数応答については、入力されたスペクトル包絡と細胞群の周波数応答特性とが一致する場合、その細胞群は強い反応を示すことが生理学的実

験からの知見として報告されている。すなわち、細胞群の周波数応答はそれぞれの細胞群における周波数応答特性と入力スペクトルのマッチングの割合によって決定される。以上のことから、スペクトル表現モデルでは入力スペクトル $p(x)$ と各細胞の周波数応答関数 $\psi(b, a, \phi)$ の内積によって各細胞の周波数応答 $r(b, a, \phi)$ を定義している。

$$r(a, b, \phi) = \psi(a, b, \phi) * p(x; a, b)$$

3.2.2 時間応答特性

特徴周波数に近い刺激を与えた場合、神経細胞群は刺激の立ち上がりでのみ強く反応し、その後は時間とともに反応の振幅と帯域幅がともに減衰していく。スペクトル表現モデルでは、この時間応答特性を指数減衰として式 (3.1) のような関数によって表す。

$$\lambda(t) = \exp\left(\frac{-t}{\tau}\right) \quad (3.1)$$

ここで、 τ は時定数である。

各細胞群の周波数応答に対して式 (3.1) に示す時間応答関数を畳み込むことにより、時間的な特性を含んだ時間-周波数応答 $r(b, a, \phi, t)$ は次式のように表される。

$$\begin{aligned} r(a, b, \phi, t) &= r(a, b, \phi) * \lambda(t) \\ &= \psi(a, b, \phi) * p(x; a, b, t) * \lambda(t) \end{aligned}$$

さらに、任意の時刻における神経細胞群全体の応答は、各細胞群の時間-周波数応答をそれぞれの時刻において加算することにより求められる。

3.3 スペクトル表現モデルにおける処理の特徴

蝸牛においては音響信号に対して周波数分析を行ない、対数周波数軸上へのマッピングを行なっている。一方、1次聴覚野においてはスペクトル包絡の変動周波数を対数スケール軸上へマッピングしている。入力である音響信号をスペクトル包絡、蝸牛フィルターを

応答関数を用いて表現する定 Q フィルタと見なすことにより、1次聴覚野における処理と蝸牛における処理は機能的に類似した処理であると考えることが可能である。

1次聴覚野における基本的な処理は、特徴周波数軸上に存在し帯域幅と対称性で定義された定 Q フィルタにより構成されるフィルタバンクを用いた、入力スペクトルに対する周波数方向への局所フーリエ変換、もしくはウェーブレット変換と同等なものと考えることができる。また、各神経細胞群の応答は応答関数の帯域幅と対称性を変化させ入力との局所的なマッチングの度合を取っていることから、応答関数を基底関数とした入力スペクトルの多重度解析を行なっていると考えることができる。このスペクトル表現では帯域幅、すなわちスケールを変化させることにより、スペクトル包絡の大まかな形状から局所的な詳細までを表現可能である。

以上のことから、蝸牛と1次聴覚野の連続処理は音響信号に対する2回のフーリエ変換もしくはウェーブレット変換と考えることができる。また、この2回の変換処理はスペクトル包絡に対するケプストラム分析と同等の処理であると思なすこともできる。しかし、1次聴覚野スペクトル表現モデルが周波数軸に沿った局所的なスペクトル構造を扱っているのに対し、ケプストラム分析は、例えば低次のケプストラム係数がスペクトル包絡の全体の概形と傾きを表すように、スペクトル包絡の包括的な領域を対象としている。従って、スペクトル包絡の情報は全てのケプストラム係数に含まれており、それぞれを独立に扱うことはできない。これに対し、1次聴覚野スペクトル表現モデルにおいては、それぞれのスケールにおいて局所的なスペクトル包絡の情報が含まれており、それぞれの情報を独立して扱うことができる。

第 4 章

周波数変化予測追跡モデル

前述の通り、本モデルにおけるスペクトル表現は Wang と Shamma による 1 次聴覚野スペクトル表現モデルを採用する。これによりスペクトル包絡の局所的な情報を、振幅、周波数、帯域幅、位相の 4 パラメータにより表現することが可能である。本研究ではスペクトルピークに対する予測追跡が目的である。よって、それぞれのピークにおける情報を抽出し各パラメータに対して予測追跡を行なうには、スペクトル包絡の局所的な情報を独立に扱うことのできる 1 次聴覚野におけるスペクトル表現が適している。

周波数変化予測追跡モデルの流れを Figure 4.1 に示す。

モデルへの入力には周波数単位を Hz から ERB rate に変換したスペクトルの時系列である。任意の時刻におけるスペクトル包絡に対して、1 次聴覚野における神経細胞群の周波数応答を近似した関数を基本ウェーブレットとしたウェーブレット分析を行ない、振幅、周波数、帯域幅、位相の 4 パラメータによる時間的に静的なスペクトル表現を求める。その情報表現における振幅の極大ピークを検出し、そのピークのパラメータを抽出する。そして各ピークにおけるパラメータの時間的変化を 2 次系のシステムによって実現した機能により予測追跡を行ない、その結果として予測されたピーク情報を用いて、任意の時刻におけるスペクトル包絡を再構成する。さらに、各時刻において再構成したスペクトル包絡に対し神経細胞群の時間応答を近似した関数によって時間方向へ拡張し、全ての時刻において加算した結果、予測追跡を考慮したスペクトルを求める。

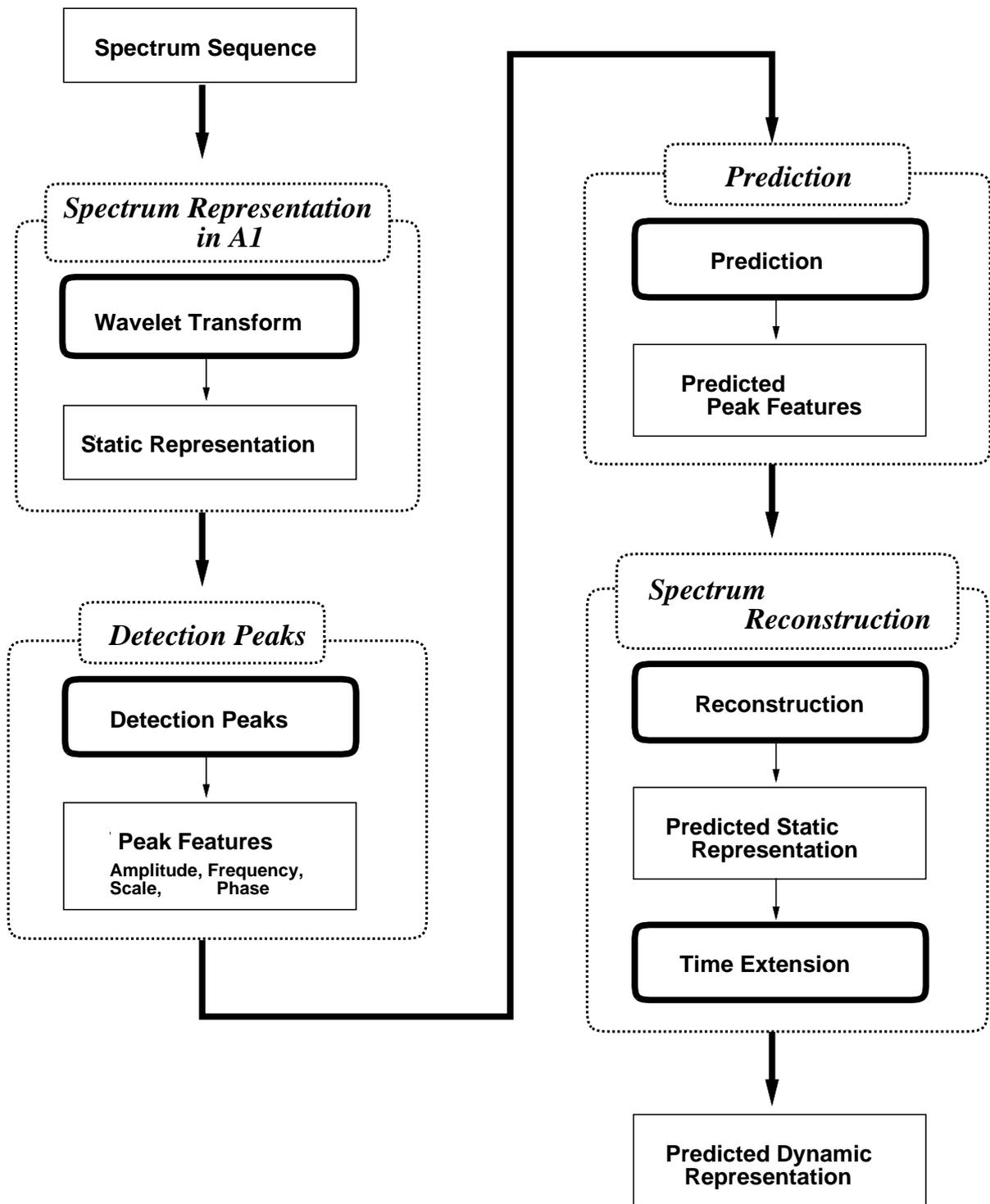


Figure 4.1: 周波数変化予測追跡モデルのフローチャート

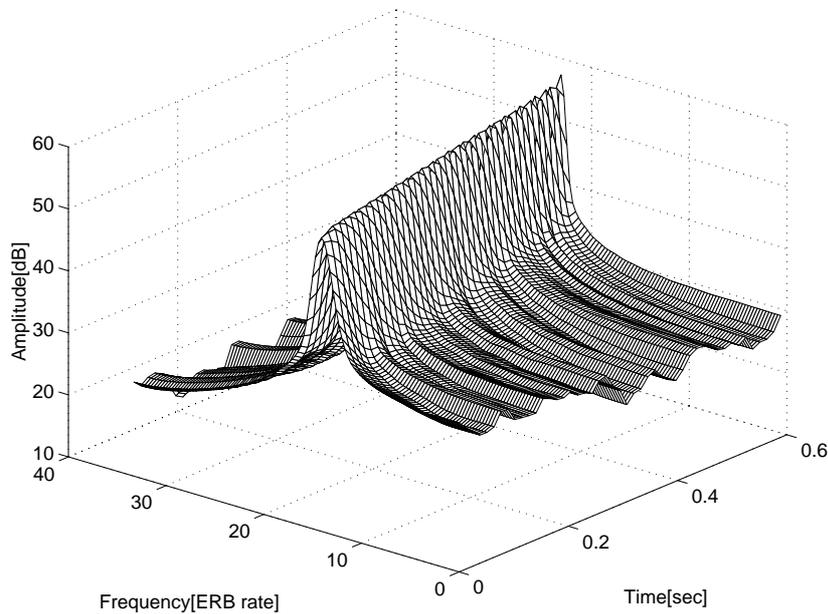


Figure 4.2: 入力スペクトル例

入力音響信号：1kHz から 3kHz までを 600msec で対数変化する上昇スイープ音

4.1 入力スペクトル

本研究では入力スペクトルを改良ケプストラム法 [15] を用いて入力音響信号より求めた。なおケプストラム係数は、次数 60 次、フレーム長 25.6msec、フレームシフト幅 6.4msec である。さらに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 7 点による移動平均処理を行なった。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学的知見と対応関係の良い ERBrate [22] に変換し表す。このスペクトル包絡を式 (4.2) に示す入力 $p(x)$ とする。

例として、1kHz から 3kHz までを 600msec で対数変化する上昇スイープ音を入力音響信号とした場合のスペクトル包絡を Figure 4.2 に示す。なおサンプリング周波数は 20kHz であり、入力音響信号には信号の立ち上がり及び立ち下がりに 0.5msec のテーパをかけた。

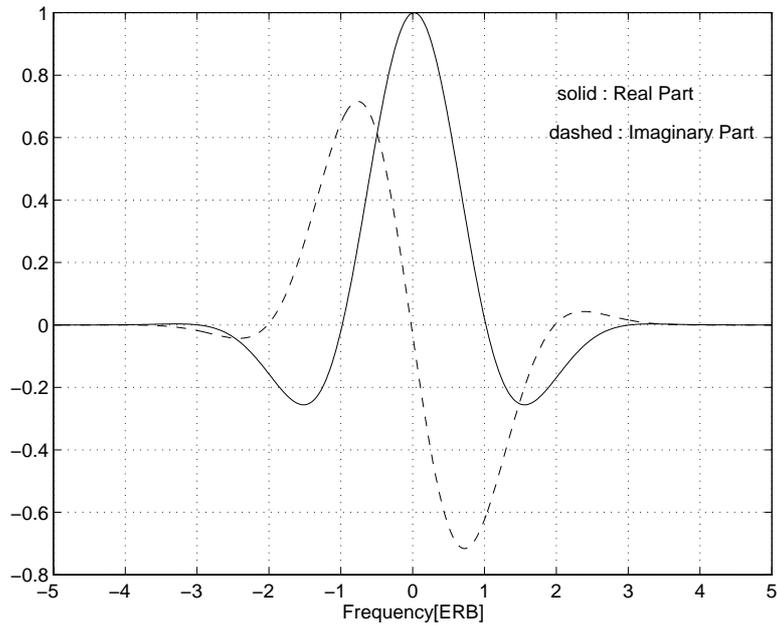


Figure 4.3: 応答関数の概形

4.2 1次聴覚野におけるスペクトル表現モデル

4.2.1 1次聴覚野細胞群における周波数応答の表現

a. 周波数応答特性の表現

細胞群の周波数応答特性に着目し、側抑制の表現を考慮した細胞群モデルの周波数応答関数を式(4.1)で表される Gabor 関数を用いて表現する。

$$\psi(x) = \exp\left(-\left(3 \cdot \frac{10^{\frac{1}{8}} - 1}{1 + 10^{\frac{1}{8}}} \cdot \omega_c x\right)^2\right) \cdot \exp(i\omega_c x) \quad (4.1)$$

ここで ω_c は Gabor 関数の中心角周波数を表す。この Gabor 関数の概形を Figure 4.3に示す。

Figure 4.3に示す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、任意の細胞群の周波数応答特性を表現することができる。

b. 周波数に対する応答

1次聴覚野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、それを内積によって表現している。ここで、任意の時刻におけるスケール a 、周波数 b における周波数応答 $r(a, b)$ は入力スペクトル $p(x)$ と周波数応答関数 $\psi(x)$ の内積を式 (4.2) のように定義することによって、 $\psi(x)$ を基本ウェーブレットとした $p(x)$ に対する周波数方向へのウェーブレット変換 [16][18] として表すことができる。

$$\begin{aligned} r(a, b) &= \langle p(x), \psi(x; a, b) \rangle_x \\ &= \frac{1}{|a|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \psi^*\left(\frac{x-b}{a}\right) p(x) dx \end{aligned} \quad (4.2)$$

ここで、 a はスケール変換のためのパラメータ、 b はシフト変換のためのパラメータであり、それぞれスケール方向、周波数方向の位置を表す。また*は復素共役を表す。式 (4.2) は、 a によって帯域幅、 b によって特徴周波数を変化させることにより表現した各細胞群の周波数応答関数 $\psi(x)$ と、入力スペクトルとのマッチングの度合として r を出力する。複素数である r より振幅と位相の値を計算することにより、任意のピークを振幅、周波数、帯域幅、位相の4パラメータで表現することができる。

Figure 4.2における時刻 300msec におけるスペクトル包絡、およびそれを式 (4.2) により周波数方向へウェーブレット分析した結果として求まる時間的に静的なスペクトル表現を Figure 4.4に示す。スケール方向の分析範囲は 0.075 ~ 0.3cycle/ERB である。

ウェーブレット分析において、スケールは分析における解像度と解釈することができ、スケールの値が大きくなるに従い解像度は細くなりスペクトルの小さな変動を分析できる。以上のことより、スケール軸は1次聴覚野スペクトル表現モデルにおけるピークの帯域幅に対応する座標軸と見なすことができる。

Figure 4.4において、周波数 20ERB rate、スケール 1.5cycle/ERB 付近が最も強く反応あり、この局所的ピークの振幅、周波数、帯域幅、位相の情報によってスペクトルにおけるピークの周波数的な位置と形状を表現することができる。

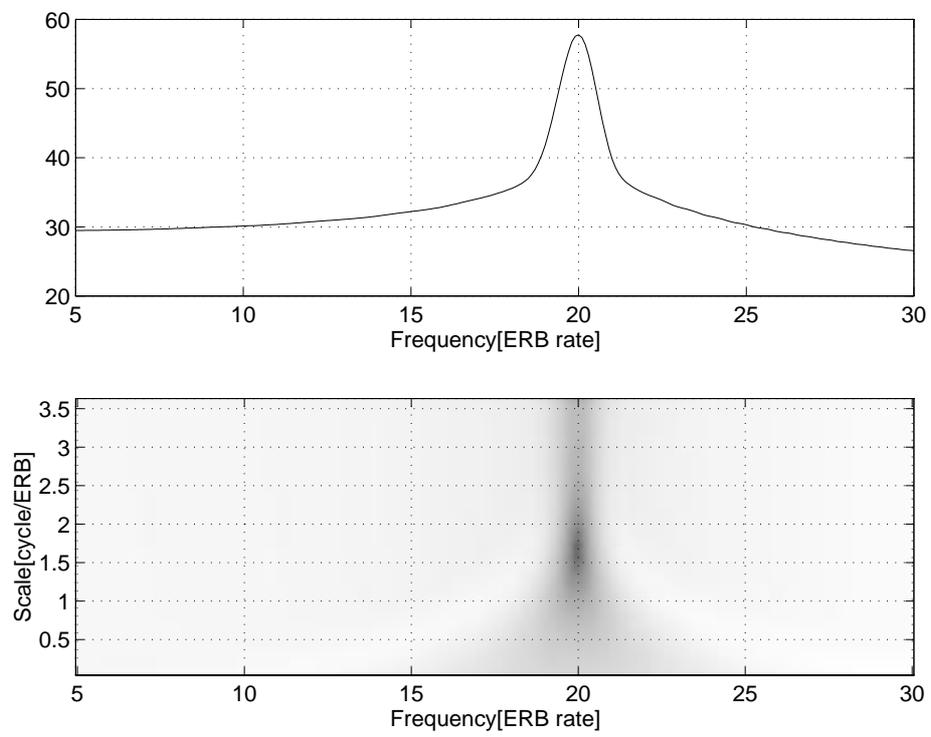


Figure 4.4: 入力スペクトル包絡とその静的スペクトル表現

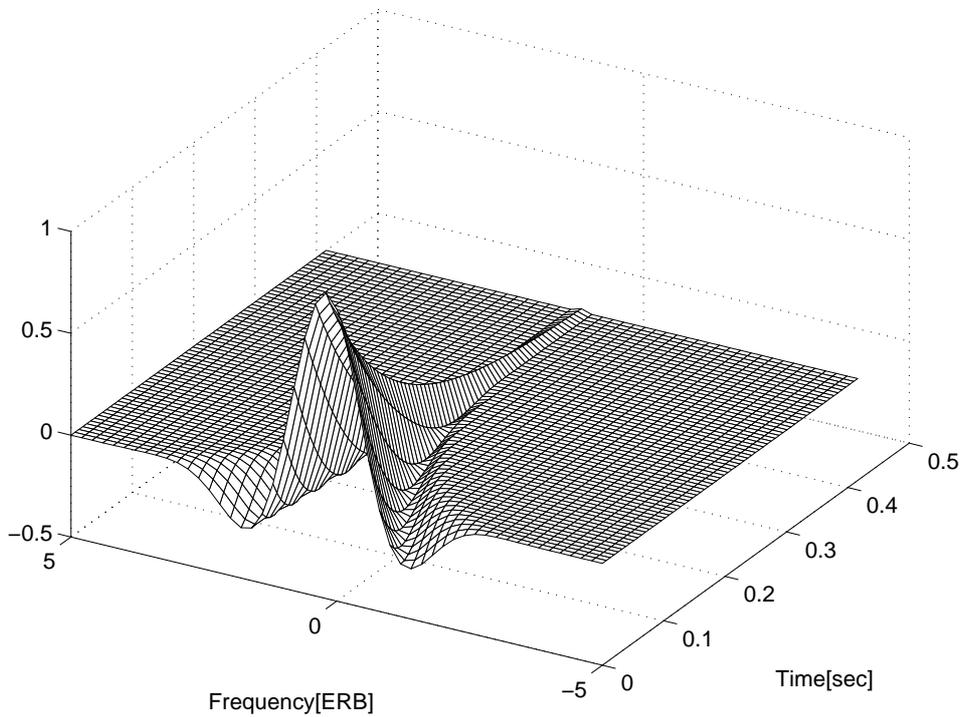


Figure 4.5: 時間-周波数応答関数の概形

4.2.2 1次聴覚野細胞群における時間応答の表現

a. 時間応答特性の表現

任意の細胞群における時間応答特性は反応の振幅 r と帯域幅 a が指数減衰するものとし、式 (3.1) に示す時間応答関数 $\lambda(t)$ をその 2 パラメータに対して適用した。

$$r(a, b, t) = r(a, b) * u(t) * \lambda(t)$$

$$a = a * u(t) * \lambda(t)$$

Figure 4.3 に示す周波数応答特性の実数部に式 (3.1) に表す時間応答特性を持たせた時間-周波数応答関数の概形を Figure 4.5 に示す。

b. 時間的变化に対する応答

時間的に静的なスペクトル表現より抽出した任意のピークにおけるパラメータを用いてスペクトルを再構成し、それに対して Figure 4.5に示す時間-周波数応答特性を畳み込むことにより、任意の時刻における時間特性を考慮した細胞群の応答を近似する。そして各時刻における応答を時間方向に加算することにより1次聴覚野における動的なスペクトル表現が得られる。ただし、特定の細胞群に対し連続して興奮刺激が入力され刺激強度がある一定値を越えた場合、その細胞群の応答は飽和状態となり、また刺激強度がある一定値以下の場合には応答しないことが生理学的にわかっている。よって、応答の時間方向への加算において応答の振幅に上限及び下限を設定した。

4.3 ピーク情報によるスペクトル表現

ピーク情報によるスペクトルの再構成について、スペクトル全体の情報を用いるウェーブレット逆変換による再構成と比較し検討する。

Figure 4.4に示すウェーブレット分析の結果を用いて、それぞれの手法により再構成したスペクトルを Figure 4.6に示す。上図がウェーブレット逆変換により再構成したスペクトルであり、下図が Figure 4.4より抽出した振幅、周波数、帯域幅、位相のピーク情報を用いて再構成したスペクトルである。なお、図中の実線は再構成したスペクトル包絡、破線は原スペクトル包絡を表し、一点破線は原スペクトルと再構成スペクトルの各周波数における差を表す。

Figure 4.6より、ピーク情報のみを用いた場合においてもピークおよ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可能であることが確認できる。

本研究では、4.2節において述べた1次聴覚野スペクトル表現モデルを用いて求めた振幅、周波数、スケール(帯域幅)、位相の各軸により表されるスペクトル表現より、局所的ピークにおけるそれぞれのパラメータを抽出し、予測追跡処理を行なった後、4パラメータを用いて Figure 4.3に示す応答関数を変形させ各ピークを再構成する。

音声の認識において、スペクトル包絡におけるピークの周波数、帯域幅などの情報は重要である。本研究における手法では、各ピークを4パラメータにより表現することにより、スペクトルを処理する場合において、スペクトル表現に存在するいくつかのピークの情報のみを用いてスペクトル包絡のピークおよびその付近の周波数帯域における概形を

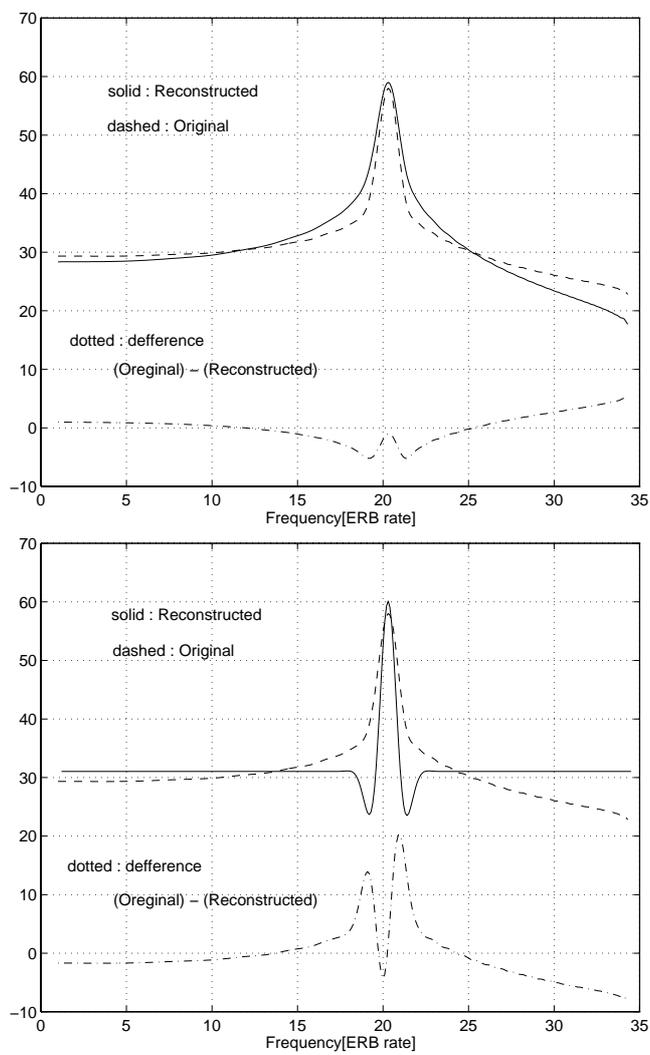


Figure 4.6: スペクトル再構成結果

上図：ウェーブレット逆変換、下図：ピーク情報使用

実線：再構成スペクトル、破線：原スペクトル、一点破線：両スペクトルの差

表現することが可能である。このことよりスペクトル表現において、スペクトル全体の情報を用いる手法と比較し、本モデルにおいて用いたピーク情報抽出による表現手法は、ピーク情報に着目した情報圧縮を行なっていると解釈できる。

4.4 周波数変化に対する予測追跡機能

4.4.1 予測追跡機能の定式化

本モデルにおける予測追跡機能は、2次系のシステムによって実現する。

河原 [17] により、基本周波数追跡制御を行なう聴覚フィードバック系は早い応答と遅い応答を示す2種類の2次系の和として良く近似できることが報告されている。この聴覚フィードバック系は知覚から発声までのピッチ制御特性を含んでおり、聴覚におけるスィープ音やホルマントなどのスペクトルピークに対する追跡機能と関係があるものと考えられる。

2次系は、その解を $p, q = c \pm jd$ とすると式 (4.3) となる。

$$\begin{aligned} H(s) &= \frac{pq}{(s+p)(s+q)} \\ &= \frac{c^2 + d^2}{s^2 + 2cs + (c^2 + d^2)} \end{aligned} \quad (4.3)$$

ここで、 $\omega^2 = c^2 + d^2$ とすると、

$$\begin{aligned} d &= \sqrt{\omega^2 - c^2} \\ &= \sqrt{1 - \frac{c^2}{\omega^2}} \cdot \omega \end{aligned}$$

また、 $\zeta = \frac{c}{\omega}$ とすると、

$$c = \zeta\omega$$

よって、標準的な伝達関数 $H(s)$ である式 (4.4) が導かれる。

$$H(s) = \frac{\omega^2}{s^2 + 2\zeta\omega s + \omega^2} \quad (4.4)$$

ここで、 ζ 、 ω はそれぞれ2次系の減衰定数、固有角周波数を表す。また $\zeta = 1$ の場合、臨界制動2次系となる。

ζ は式(3.1)に与えられた時定数 τ を用いて、式(4.5)のように求めることができる。

$$\zeta = \frac{1}{\tau\omega} \quad (4.5)$$

式(4.4)をインパルス不変法を用いて $H(s)$ - $H(z)$ 変換を行なうと、式(4.6)のように表される。

$$H(z) = \frac{Gz^{-1}}{1 - 2\exp^{-\zeta\omega T} \cos(\sqrt{1 - \zeta^2} \cdot \omega T)z^{-1} + \exp^{-2\zeta\omega T} z^{-2}} \quad (4.6)$$

G はゲイン定数、 T はサンプリングの時間間隔である。ここで、固有周波数を f_n 、サンプリング周波数を f_s を用いて $-\pi$ から π に正規化した周波数 λ を次のように表す。

$$\begin{aligned} \omega &= 2\pi f_n \\ T &= \frac{1}{f_s} \\ \lambda &= \omega T = 2\pi \frac{f_n}{f_s} \end{aligned}$$

よって、 $0 < \zeta < 1$ の場合、式(4.6)は次のようなAR(Auto Regressive)モデルで表される。

$$\begin{aligned} H(z) &= \frac{Gz^{-1}}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}} \quad (4.7) \\ \alpha_1 &= -2\exp^{-\zeta\lambda} \cos(\sqrt{1 - \zeta^2} \cdot \lambda) \\ \alpha_2 &= \exp^{-2\zeta\lambda} \end{aligned}$$

α_1, α_2 は線形予測係数である。

入出力系列の z 変換 $X(z), Y(z)$ および伝達関数 $H(z)$ の関係は $Y(z) = H(z)X(z)$ となることより、式(4.8)の関係式が得られる。

$$Y(z)[1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}] = X(z)[Gz^{-1}] \quad (4.8)$$

上式を逆 z 変換すると、時刻 t の関数として入力 $x(t)$ 、出力 $y(t)$ の時系列の関係は式(4.9)となる。

$$\begin{aligned} y(t) + \alpha_1 y(t-1) + \alpha_2 y(t-2) &= Gx(t-1) \\ y(t) &= Gx(t-1) - \alpha_1 y(t-1) - \alpha_2 y(t-2) \end{aligned} \quad (4.9)$$

式(4.9)のように、任意の時刻 t における出力 $y(t)$ はその時刻の 1 時刻前の入力 $x(t-1)$ と、1 時刻および 2 時刻前の出力 $y(t-1), y(t-2)$ の線形結合によって求めることができる。

式(4.9)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必要がある。よって G は次式によって得られる。

$$G = 1 + \alpha_1 + \alpha_2$$

4.4.2 予測追跡機能の特性

式(4.9)において、右辺の第 2 項と第 3 項は 1 時刻および 2 時刻前の出力 $y(t-1), y(t-2)$ を $y(t-1)$ 方向に外挿する慣性を表す。また $-\alpha_1 - \alpha_2 = 1 - G > 0, G > 0$ より、任意の時刻における出力 $y(t)$ は 1 時刻前の入力 $x(t-1)$ と 1、2 時刻前の出力 $y(t-1), y(t-2)$ との内分点となる。以上のことから、2 次系の予測追跡機能には過去 2 時刻における出力の外分と現在の入力との内分により求めることができる。

2 次系予測追跡機能を用いて周波数変化を追跡した結果を Figure 4.7 に示す。入力信号は 1.732kHz に停留部を含み 1kHz から 3kHz までを対数変化する上昇スイープ音であり破線で示す。実線は $f_n = 5, \tau = 0.35 (\zeta = 0.4961, \alpha_1 = -1.7826, \alpha_2 = 0.8191)$ の場合、一点破線は $f_n = 7, \tau = 0.7 (\zeta = 0.2, \alpha_1 = -1.8191, \alpha_2 = 0.8935)$ の場合における予測追跡の軌跡である。

本研究においては、静的なスペクトル表現から抽出したピークの 4 パラメータすべてに対して 2 次系予測追跡機能を適用し、軌跡の予測追跡を行なった。

4.4.3 雑音が付加された場合の予測

追跡対象信号の途中に、もしくは後続する形で雑音が付加された場合の予測追跡処理について述べる。

雑音入力された場合、その時刻における対象信号のピーク情報が抽出できない。そのため、雑音入力中は式(4.9)における 1 時刻前の入力 $x(t-1)$ の代わりに 1 時刻前の出力 $y(t-1)$ を用いて予測追跡を行なう。これにより雑音中を外挿する形での予測追跡が表現できる。

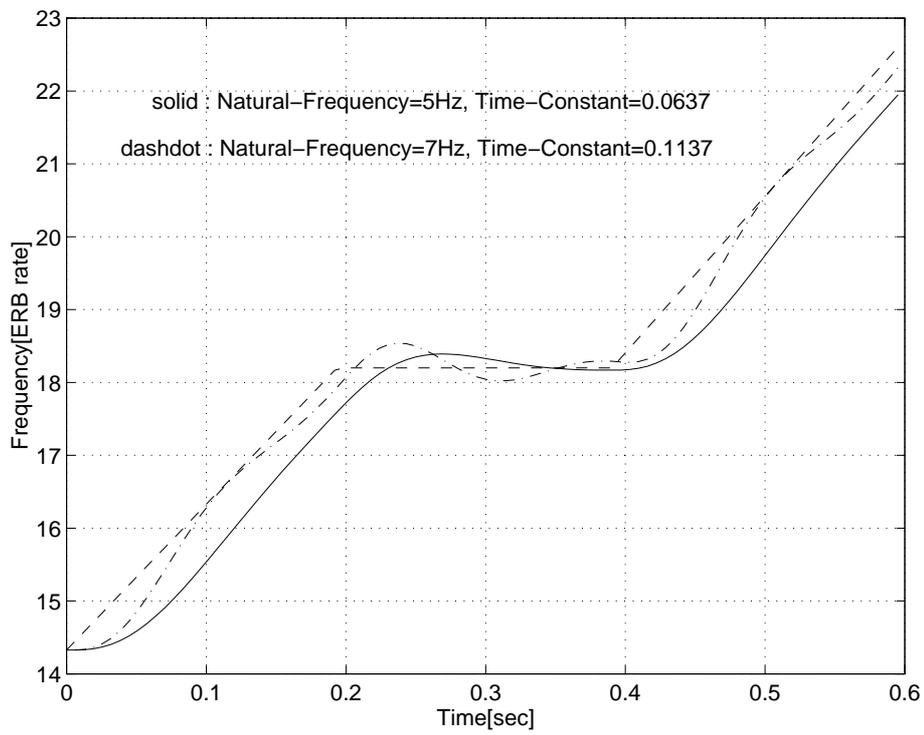


Figure 4.7: 2次予測追跡機能の特性

入力信号 : 1.732kHz に停留部を含み 1kHz から 3kHz までを対数変化する上昇スイープ音

第 5 章

周波数変化予測追跡シミュレーション

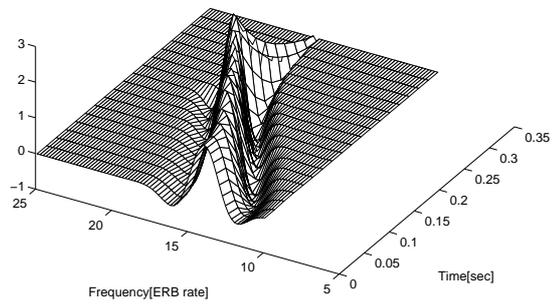
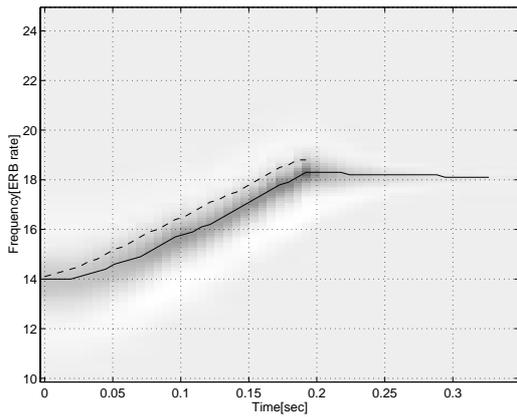
先述べたように、音声の認識にはホルマント軌跡の予測追跡が重要である。本研究ではその基礎実験として、単一のスペクトルピークが存在する対数線形スイープ音を対象として予測追跡シミュレーションを行なった。以下の節において、対数線形スイープ音、段つき対数線形スイープ音、雑音による中断のある対数線形スイープ音、雑音を後続させた対数線形スイープ音、および連続発話における母音のホルマント遷移を模倣した複数スイープ音に対するシミュレーションについて述べる。

5.1 対数線形スイープ音に対する予測追跡結果

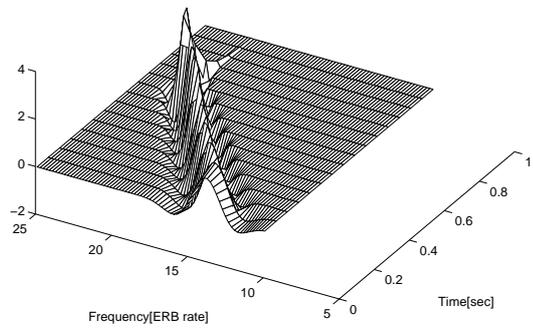
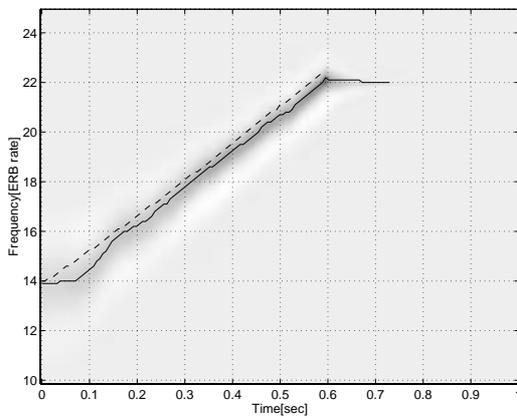
1kHz から 2kHz の帯域を 200msec で上昇する対数線形スイープ音、および 1kHz から 3kHz の帯域を 600msec で上昇する対数線形スイープ音を入力音響信号としてとしてシミュレーションを行なった結果を Figure 5.1 に示す。なお時定数 $\tau = 63.7\text{msec}$ (減衰定数 $\zeta = 0.5$)、固有振動数 $f_n = 5\text{Hz}$ である。Figure 5.1 において実線はスペクトル表現のピークの軌跡、破線は入力した対数線形スイープ音の軌跡を表す。

Figure 5.1 より、スペクトル表現は始端部において変化が穏やかであり、徐々に変化の割合が増加している。その結果、スイープ音の変化が遅れる方向に湾曲しており、対数線形スイープ音に対する心理物理実験の結果とほぼ同様な結果が得られた。

また、1kHz から 3kHz の帯域を 600msec で変化する場合の周波数変化率は 3.333kHz/sec 、1kHz から 2kHz の帯域を 200msec で変化する場合は 5.0kHz/sec である。Figure 5.1 に示すように、周波数変化率が大きい程追跡軌跡の遅れが大きくなることがわかる。

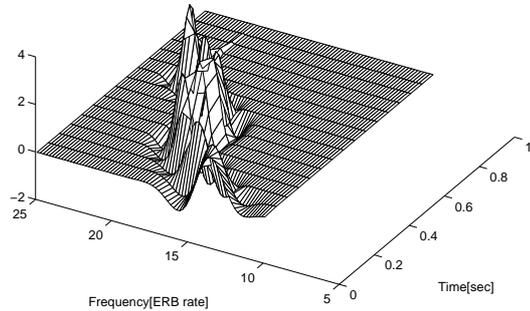
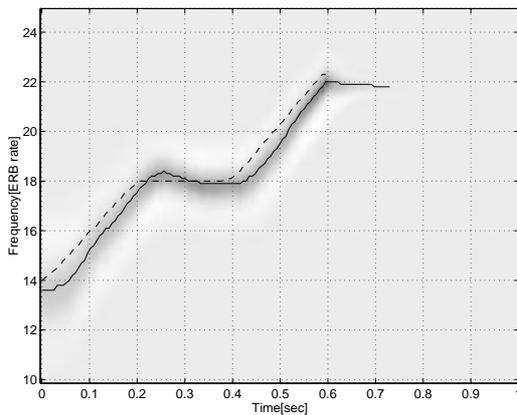


(a) Frequency transition:1kHz-2kHz, Duration:200msec

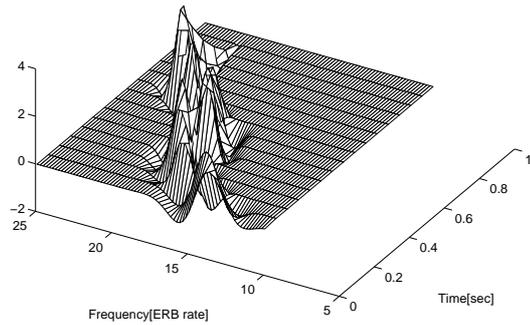
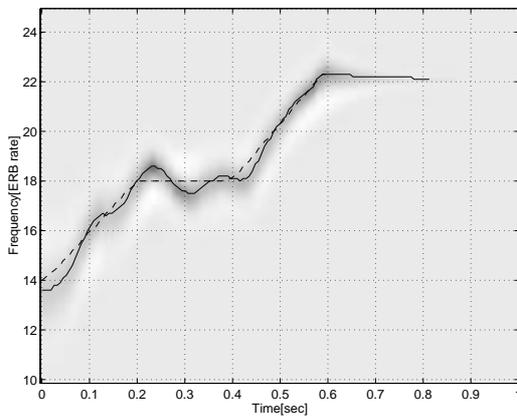


(b) Frequency transition:1kHz-3kHz, Duration:600msec

Figure 5.1: 対数線形スweep音に対する予測追跡



(a) Frequency transition:1kHz-1.732kHz-3kHz, Duration:600msec



(a) Frequency transition:1kHz-1.732kHz-3kHz, Duration:600msec

Figure 5.2: 段つき対数線形スイープ音に対する予測追跡

5.2 段つき対数線形スイープ音に対するシミュレーション

1kHz から 3kHz の帯域を上昇する対数線形スイープ音に対して、 $\sqrt{3} = 1.732\text{kHz}$ に停留部を含む段つきのスイープ音を入力音響信号としたシミュレーション結果を Figure5.2 に示す。 $\tau = 63.7\text{msec}$ ($\zeta = 0.5$)、 $f_n = 5\text{Hz}$ の場合 (a) と、 $\tau = 113.7\text{msec}$ ($\zeta = 0.2$)、 $f_n = 7\text{Hz}$ の場合 (b) である。Figure5.2において実線はスペクトル表現のピークの軌跡、破線は入力した対数線形スイープ音の軌跡を表す。

Figure5.2において、相川らによって報告された停留部におけるピッチの揺らぎが発生していることを確認できる。また、固有振動数と時定数を変化させることにより、ピッチ

の揺らぎの程度や回数を制御することができる。

5.3 雑音を後続させた対数線形スイープ音に対するシミュレーション

1k ~ 2kHz の周波数帯域を 200msec で上昇する対数線形スイープ音、および 1k ~ 3kHz の周波数帯域を 600msec で上昇する対数線形スイープ音に対しそれぞれ 200msec の白色雑音を後続させた音響信号に対して予測追跡シミュレーションを行なった。 $\tau = 99.5\text{msec}$ ($\zeta = 0.2$) $f_n = 8\text{Hz}$ の場合のシミュレーション結果を Figure 5.3 に示す。Figure 5.3 において実線はスペクトル表現のピークの軌跡、破線は入力した対数線形スイープ音の軌跡を表す。

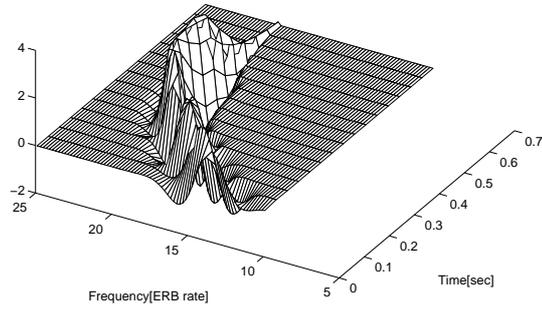
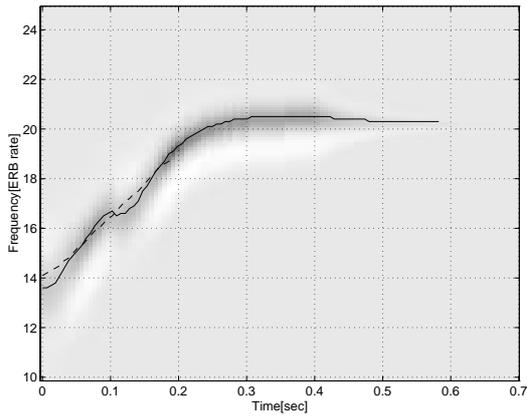
Figure 5.1 のようにある時刻において入力が途切れた場合と比較して、Figure 5.3 のようにスイープ音の後に雑音が付加された場合には雑音中を引き続き予測追跡していく。これは倉片らの心理物理実験より示唆された、スイープ音が後続雑音中を外挿して知覚される現象と同様な結果である。

5.4 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

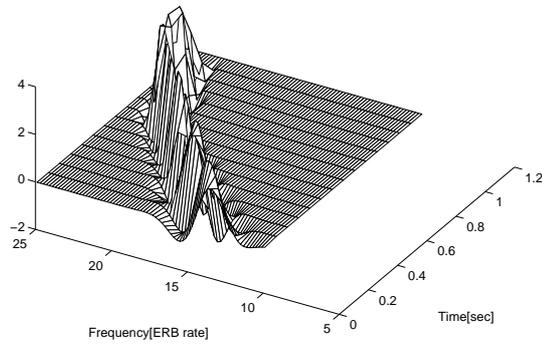
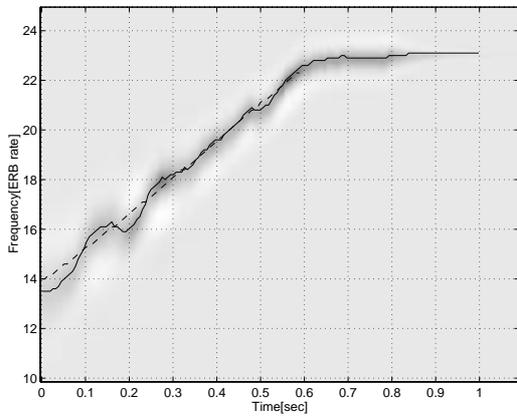
対数線形的に変化するスイープ音の後に白色雑音を付加し、雑音終了後に対数線形スイープ音が再び出現するような音響信号を対象としてシミュレーションを行なった。

1k ~ 2kHz の周波数帯域を 300msec で上昇する対数線形スイープ音の、100 ~ 200msec の時間を白色雑音によって置換した入力信号、および 1k ~ 3kHz の周波数帯域を 600msec で上昇する対数線形スイープ音の、200 ~ 400msec の時間を白色雑音によって置換した入力信号に対する予測追跡結果を Figure 5.4 に示す。なお $\tau = 63.7\text{msec}$ ($\zeta = 0.5$) $f_n = 5\text{Hz}$ であり、実線はスペクトル表現のピークの軌跡、破線は入力したスイープ音の軌跡を表す。

Figure 5.4 において、1k ~ 2kHz の周波数帯域を 300msec で上昇する対数線形スイープ音の場合、その雑音による中断時間が 100msec と短いため、雑音中を内挿するように予測追跡していくことがわかる。しかし、1k ~ 3kHz の周波数帯域を 600msec で上昇する対

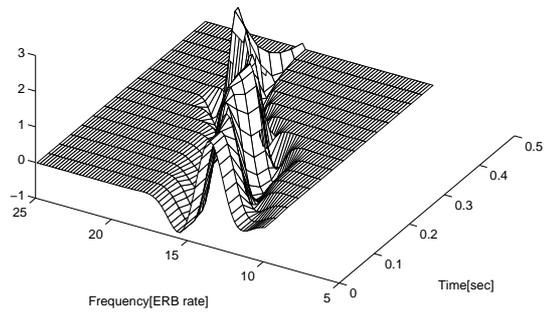
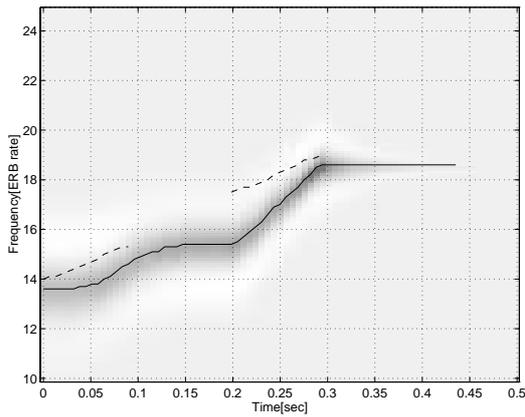


(a) Frequency transition:1kHz-2kHz, Duration:200msec

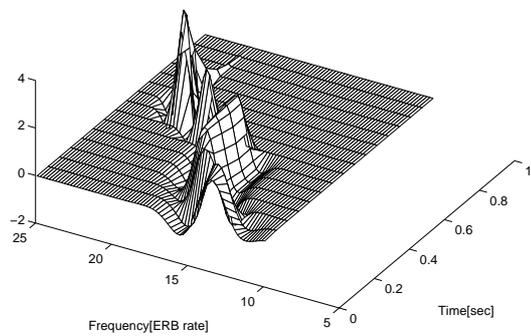
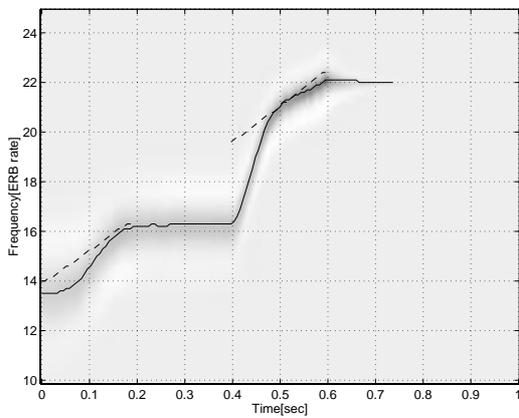


(a) Frequency transition:1kHz-3kHz, Duration:600msec

Figure 5.3: 雑音を後続させた対数線形スweep音に対する予測追跡



(a) Frequency transition:1kHz-2kHz, Duration:300msec



(a) Frequency transition:1kHz-3kHz, Duration:600msec

Figure 5.4: 雑音による中断のある対数線形スイープ音に対する予測追跡 1

数線形スイープ音に対するシミュレーションでは、雑音による中断が 200msec であり、雑音中において予測追跡が中断していることがわかる。

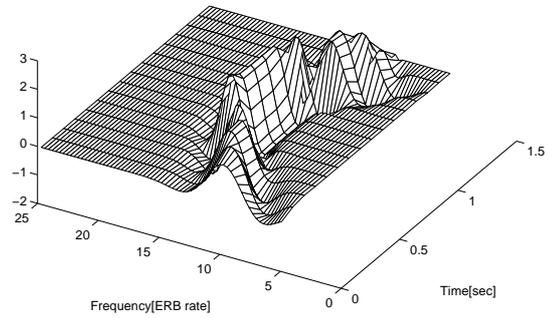
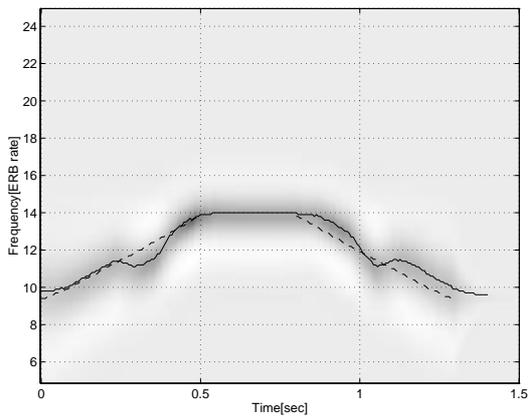
倉片らは、スイープ音の途中で雑音が付加された場合における雑音中の予測追跡について、雑音入力開始後約 100msec は外挿的な予測を継続することを報告している。しかし同時に、雑音による中断が 400msec の場合においても雑音中にスイープ音を知覚しているとも報告している。しかしこれはスイープ音の再入力を知覚した後、その時刻以降の情報をを用いた時間的に逆向きの予測が働くためと考えられる。本モデルではこのようなトップダウン的処理は考慮しておらず、中断時間が長い場合の予測追跡は不可能である。

また倉片らの心理物理実験と同様の入力信号に対してシミュレーションを行なった。入力信号は、500 ~ 1kHz を 500msec で上昇する対数線形スイープ音の後に 300msec の白色雑音を付加し、さらにそれに後続させる形で持続時間が 500msec の、1k ~ 500Hz で下降、1k ~ 2kHz で上昇、1.5157k ~ 3.0314kHz で上昇するスイープ音を付加したものである。Figure 2.2 に示した知覚パターンを示す予測追跡シミュレーション結果およびそれぞれの場合におけるパラメータを Figure 5.5 に示す。

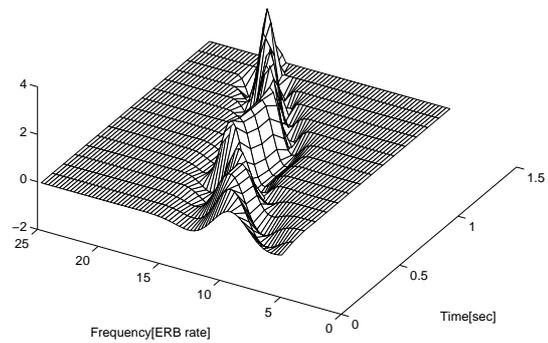
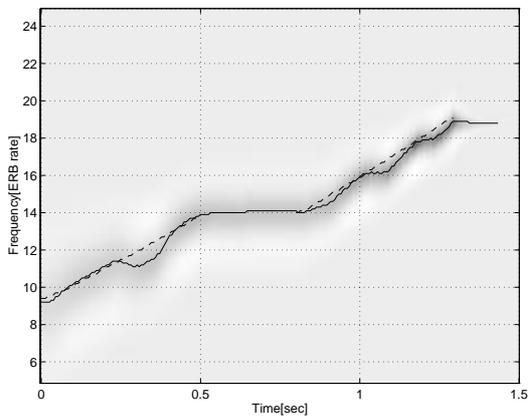
固有振動数及び時定数を変化させることにより、心理物理実験において知覚されている軌跡を周波数変化音予測追跡モデルによって再現することが可能である。

5.5 連続発話における母音のホルマント遷移を模倣した複数スイープ音に対する予測追跡結果

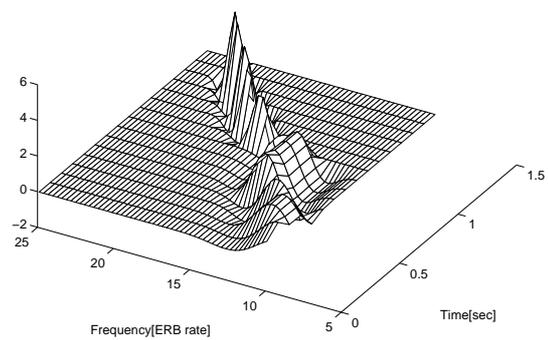
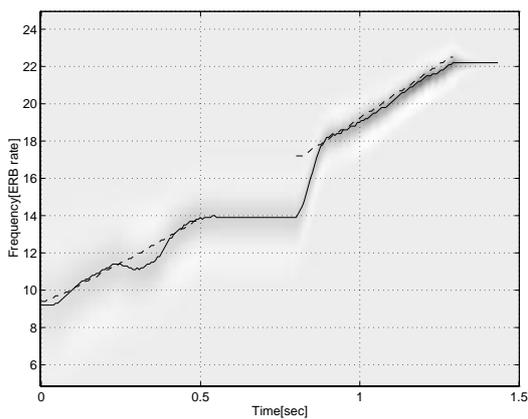
母音を連続的に発話した場合の各母音を特徴付けるホルマント周波数が前後の母音の影響を受けなまける現象、すなわち調音結合に対する基礎的なシミュレーションを行なった。シミュレーションに用いた刺激音は、3 連続母音 /iai/ の第 1 から第 3 ホルマントまでの周波数遷移をそれぞれスイープ音で模倣した複合音である。それぞれの母音における第 1 から第 3 ホルマントの周波数は、/i/ においては 357Hz、2157Hz、3043Hz であり、/a/ においては 843Hz、1200Hz、2543Hz である??。また、母音定常部の長さは 120msec、各母音間の遷移部の長さは 30msec であり、遷移は対数線形的に変化させた。この疑似ホルマントを対象としたシミュレーションを $\tau = 63.7\text{msec}$ ($\zeta = 0.5$) $f_n = 6\text{Hz(a)}$ と、 $\tau = 99.5\text{msec}$ ($\zeta = 0.2$) $f_n = 8\text{Hz}$ のそれぞれの場合について行なった。なお、ホルマントに対する予測追跡はそれぞれのホルマント遷移の時系列に対して独立に行ない、スペクトルの再構成



(a) Frequency transition:500Hz-1kHz-500Hz, Duration:1.3sec

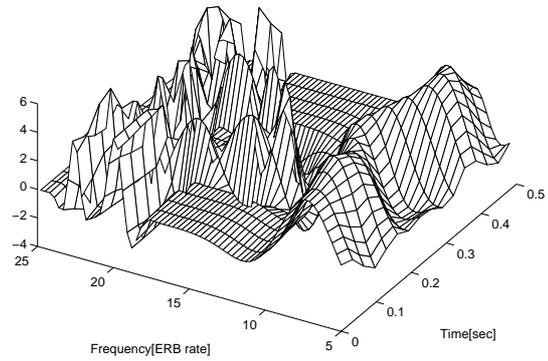
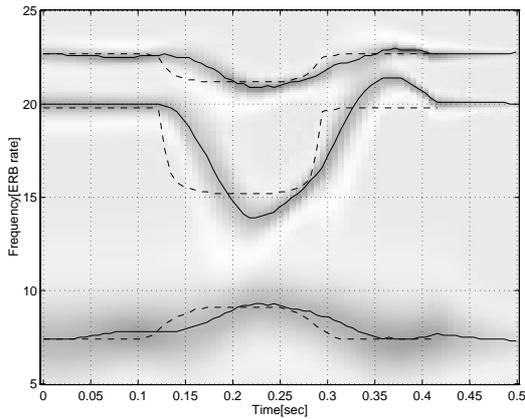


(b) Frequency transition:500Hz-1kHz-2kHz, Duration:1.3sec

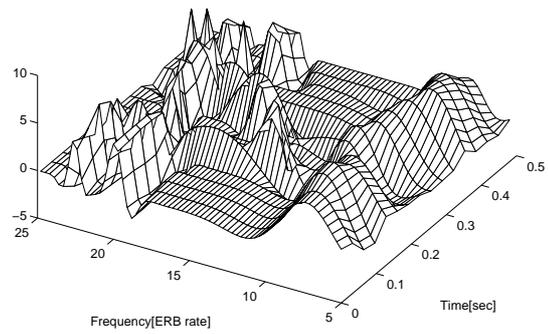
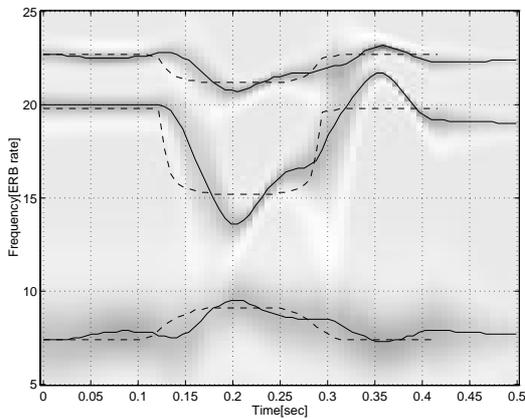


(c) Frequency transition:500Hz-1kHz-1.515kHz-3kHz, Duration:1.3sec

Figure 5.5: 雑音による中断のある対数線形スイープ音に対する予測追跡2
倉片らの心理物理実験結果を模擬したシミュレーション



(a) Imitation of continuous vowels /iai/



(b) Imitation of continuous vowels /iai/

Figure 5.6: 連続母音のホルマント遷移を模倣した複数スweep音に対する予測追跡

処理においてそれぞれの予測を考慮したホルマント遷移を統合した。

シミュレーション結果を Figure 5.6 に示す。なお、Figure 5.6 において実線はスペクトル表現のピークの軌跡、破線は入力した対数線形スweep音の軌跡を表す。

Figure 5.6 において、予測追跡した軌跡は入力軌跡に対して全体的に時間的な遅延が生じる。また、/a/ の母音中心付近において overshoot が発生しており、overshoot の持続時間や度合は予測追跡のパラメータによって制御できることが確認できる。

固有振動数及び時定数を変化させることにより、知覚レベルにおける音響特徴の補正を、周波数変化音予測追跡モデルを用いることにより信号処理レベルにおいてシミュレーションによって再現可能であると考えられる。

第 6 章

結論

本研究では、音韻修復など人間の聴覚における音情報の補完現象の実現を目的として、1 次聴覚野におけるスペクトル表現の知覚モデルに基づき周波数変化音に対する予測追跡機能をモデル化した。

本モデルではピークの予測追跡に適したスペクトル表現を得るため、Wang と Shamma によって提案された 1 次聴覚野におけるスペクトル表現モデルをより工学的手法を用いて実現した。この 1 次聴覚野のスペクトル表現モデルを用いることにより、スペクトルは周波数、帯域幅、位相の各軸で構成される空間に表現される。このスペクトル表現におけるピーク情報のみを用いてスペクトルの再構成を行なった結果、スペクトルの概形を再構成することが可能であることがわかった。このように、1 次聴覚野におけるスペクトル表現を用いることにより、スペクトルの局所的な情報を独立して扱うことが可能であり、これはスペクトル表現における情報圧縮と考えることができる。本研究は、特にスペクトルのピーク遷移に着目したことから、ピークにおける振幅、周波数、帯域幅、位相の 4 パラメータからなる情報抽出した。

また、過去の周波数変化の履歴より以後の軌跡を予想し追跡する機能は、2 次系のシステムによって実現した。この予測追跡機能を用いて、1 次聴覚野におけるスペクトル表現から抽出した一連のピーク列の変化に対して予測追跡を行なう。そして、予測したピーク列から各時刻におけるスペクトルを再構成し、それに対して神経細胞群の時間応答を近似した減衰特性を持たせ、さらに各時刻における応答を時間方向に加算することにより、予測追跡を考慮したスペクトルを求めた。

ここで、2 次系の予測追跡機能と時間減衰特性を付加した応答の加算は、予測追跡に関

してそれぞれ異なった性質を持つ。前者は目的からのずれを検出し情報の変化を予測追跡する特性、後者は過去の履歴を維持し変化を抑制する特性をそれぞれ有している。加算による変化抑制の特性は神経系における情報処理に要する時間に依存したむだ時間要素と考えることができる。しかし、予測追跡機能の存在を示唆する多くの研究結果が発表されている中において、目的を能動的に予測追跡する特性を司る機構については、現在のところその断定は不可能である。予測追跡機能のシステムを同定するために、聴覚系高次における能動制御に関する今後の研究が期待される。

本研究において構築した周波数変化音予測追跡モデルは周波数変化音の軌跡に対して外挿及び内挿を行なう機能を含んでおり、知覚レベルにおいて音脈を形成する一要因である周波数変化の連続性に関して、信号処理レベルにおける表現が実現可能である。

モデルを検証するため、様々な対数線形スイープ音を対象として予測追跡シミュレーションを行なった。その結果、モデルのパラメータを変化させることにより、対数スイープ音の湾曲知覚現象や雑音が付加されたスイープ音の補完追跡などの心理物理学実験の結果と類似した結果を得ることができ、人間の知覚の信号処理レベルにおける近似表現が可能であることを確認した。

本モデルで実現した聴覚における周波数変化音予測追跡機能を計算機に実装することにより、実環境下における音声認識技術の高精度化や計算機による聴覚的情景分析の実現などが期待できる。

謝辞

本研究を進めるにあたり、常に御助言を頂き、終始にわたり御指導頂いた赤木正人助教授に深謝致します。また、多くの有益な御助言を頂いた飯島泰蔵教授、岩城護助手に深謝致します。さらに、最後となりましたが日常生活を通し多大な御世話、御迷惑をおかけし、また非常に有意義な議論をともにして下さった飯島赤木両研究室の皆様方に心より感謝致します。

参考文献

- [1] A.S.Bregman, AUDITORY SCENE ANALYSIS, MIT press, London, 1990.
- [2] 相川 清明, 津崎 実, 周波数変化音知覚における残効, 信学技報, vol.95, SP95-42, pp.1-8, 1995.
- [3] 相川 清明, 津崎 実, 種々の周波数変化音の動的知覚特性, 信学技報, vol.95, SP95-136, pp.7-14, 1996.
- [4] 相川 清明, 津崎 実, 河原 英紀, 東倉 洋一, 周波数変化音追跡の動特性, 日本音響学会誌, vol.52, no.10, pp.741-751, 1996.
- [5] 増田 郁代, 相川 清明, 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.
- [6] 増田 郁代, 相川 清明, 津崎 実, 連続母音の過渡部の知覚における後続雑音の効果, 音響学会聴覚研究会資料, H95-61, 1995.
- [7] 古井 貞熙, デジタル音声処理, 東海大学出版会, 東京, 1985.
- [8] 片山 徹, フィードバック制御の基礎, 朝倉書店, 東京, 1987.
- [9] 城戸 健一, デジタル信号処理入門, 丸善, 東京, 1985.
- [10] 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(続報), 信学技報, vol.95, SP95-43, pp.9-16, 1995.
- [11] 松井 理直, 倉片 憲治, 西村 篤, 知覚的軌跡の形成過程における予測処理と制約, 信学技報, vol.95, SP95-44, pp.17-24, 1995.

- [12] A.van Wieringen, L.C.W.Pols, Discrimination of single and complex consonant-vowel and vowel-consonant like formant transitions, *J.Acoust.Soc.Am.*, vol.98, no.3, pp.1304-1312, 1995.
- [13] M.Akagi, Modeling of contextual effects based on spectral peak interaction, *J.Acoust.Soc.Am.*, vol.93, no.2, pp.1076-1086, 1993.
- [14] 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, 北陸先端科学技術大学院大学修士論文, 1996.
- [15] 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.
- [16] 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.
- [17] 河原 英紀, フィードバック音声の基本周波数への摂動に対する応答の成分への分析について, *音響学会聴覚研究会資料*, H95-15, 1995.
- [18] 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.
- [19] S.A.Shamma, J.W.Fleshman, P.R.Wiser, H.Versnel, Organization of Response Areas in Ferret Primary Auditory Cortex, *J.Neurophysiol.*, col.69, no.2, pp.367-383, 1993.
- [20] K.Wang, S.A.Shamma, Self-Normalization and Noise-Robustness in Early Auditory Representations, *IEEE Trans. Speech Audio Processing*, vol.2, no.3, pp.421-435, 1994.
- [21] 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.
- [22] 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.
- [23] B.C.J.Moore, 大串 健吾 訳, 聴覚心理学概論, 誠信書房, 東京, 1994.