NAIST-IS-DT0161032

博士論文

音高知覚に関する情報表現の情報理論的観点からの研究

前田 新一

2004年 3月 24日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士 (理学) 授与の要件として提出した博士論文である。

前田 新一

審査委員：　石井 信 教授
鹿野 清宏 教授
関 浩之 教授
柴田 智広 助教授

音高知覚に関する情報表現の情報理論的観点からの研究*

前田 新一

内容梗概

本論文では、脳の初期知覚系、特に音高知覚 (ピッチ) に関する情報表現について議論する。聴覚で処理される音声情報は、情報理論における情報量の観点からは非常に冗長な情報となっている。冗長性と予測可能性は密接な関係を持っているが、脳は情報の予測可能性を積極的に利用していることが認知心理学的な研究からわかっている。その典型的な例は錯覚現象に現われており、いくつかの錯覚現象は、情報の傾向を予測した推定を行っていると考えることで説明することができる。

ここではまず、エントロピー最大化に基づく聴覚モデルによって音高知覚でみられる錯覚現象を説明するとともに、いくつかの生理学的知見を説明する。このモデルにおけるエントロピー最大化は、正確には結合エントロピーの最大化を意味するが、結合エントロピー最大化を周辺エントロピーの最大化と独立化とに 2 段階にわたる手続きに分離して行う。このモデルでは、厳密には結合エントロピーを最大化させることを保証することはできないが、解剖学的な構造との対応がとりやすいモデルとなっている。

次に、この 2 段階の手続きに分離していたエントロピー最大化に基づくモデルをノイズ付き非線形独立成分分析として定式化することで理論的に一貫したモデルとした。導出したノイズ付き非線形独立成分分析は、線形なノイズ付き独立成分分析を一般化するモデルとなった。このモデルを音声、音楽データを学習させてより詳細な音高知覚シミュレーションを行った。

最後に、エントロピー最大化や独立化がなぜ効率的な情報表現と関係するかをレート歪み理論に基づく圧縮符号の観点から議論する。最適なブロック符号化による圧縮符号は、ベクトル量子化によって表現することができ、そのベクトル量子化を最適化する方

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 博士論文, NAIST-IS-DT0161032, 2004 年 3 月 24 日.

法も考案されている。しかし、ベクトル量子化は復号や最適化がベクトルの次元に対して指数的に計算量が増えてしまう問題を持っている。この問題を Product Code による構造化によって解決するとともに、その最適化法を提案する。また、Product Code による最適な圧縮符号の生成とエントロピー最大化問題がどのように関係するかについて述べる。

キーワード

情報理論, 独立成分分析, 予測, 聴覚, 音高知覚

Studies on information representation of pitch from the viewpoint of information theory*

Shin-ichi Maeda

Abstract

This thesis proposes a model of sensory information representation, especially information representation of pitch from the viewpoint of information theory. Acoustic signals, which are provided to auditory system, have a plenty of redundancy. Psychophysical studies suggest that the neural system utilizes prediction actively, which is closely related to the utilization of the redundancy. Typical evidence of prediction appears in some illusions, where illusions are caused by the prediction of natural sensory information. The neural code that reduces the redundancy of the information is often called an efficient coding.

In this thesis, I exemplify not only an illusion in perceiving a pitch, but also some physiological findings, by using an auditory model based on entropy maximization. The entropy means here the joint entropy and it is maximized in a hierarchical manner consisting of two stages. This algorithm does not guarantee exact maximization of the joint entropy, but has a biologically reasonable architecture.

Next, I present a model that improves the above two-stage entropy maximization model into a nonlinear noisy independent component analysis (ICA). The derived learning rule naturally involves a linear ICA as well as a linear noisy ICA. This model is able to also simulate the psychological findings of pitch sensation.

Finally, I discuss why the entropy maximization or ICA is important for an efficient code. The code efficiency is discussed from a viewpoint of rate-distortion theory, i.e., an efficient code corresponds to a lossy compressed code. Vector quantization is known as

*Doctor's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DT0161032, March 24, 2004.

the best lossy source coding method among the block codes because of its satisfactory ability of expression. Although it can represent any block codes and has optimization methods that guarantee local optimality, the required encoding and optimization computation increases exponential to the input length. I propose an optimization method of a product code, which avoids the computational problem of vector quantization by a reasonable restriction of model architecture. The performance of the product code is evaluated by a simple problem.

Keywords:

information theory, independent component analysis, prediction, audition, pitch sensation

目 次

第1章　序論

本論文では、脳の初期知覚系の情報表現、特に音高知覚に関する情報表現について議論する。脳の情報表現の解明は、意義のある問題である。例えば、現在の機械学習の枠組みでは解決が困難な難題であるフレーム問題[1]を脳はなんらかの方法で解決しているようにみえる。そのため、この難題を解いている人間の脳の情報表現を明らかにすることでフレーム問題を解く手がかりが得られる可能性がある。

古くから初期知覚系の情報表現がどのようなものであるかについては、実験的にも理論的にも研究がされている。Hubel と Wiesel の一次視覚野における特定の向きの線分に特異的に応答するニューロンの発見 [3] 以来、ニューロンの受容野がどのようなものであるか、またそこから脳の特徴抽出がどのようなものであるかについて議論されている。しかし、実験によって知ることができるのは、ある物理刺激に対するニューロンの発火パターンに留まる。そのため、フレーム問題を解く脳の動作原理を知りたいなどの工学的な立場からは、実験的に調べられたニューロンの入出力関係を数理モデルで表すことができるだけでは不十分である。また、ニューロンの入出力関係を解析して、システムを推定することは難しい問題である。私は、光計測で得られる一次聴覚皮質の活動を独立成分分析 (ICA) を用いて解析した [4] [5][6] が、線形変換である ICA は常に何らかの解を導き出す。そのため、その解の中のどれが神経活動でどれがノイズであるかを判断することさえ困難であったし、本当に独立な活動をしているといえるのかどうかもそれだけでは判断がつかなかった。脳を含めて複雑な未知のシステムの解析は、様々な可能性を含む逆問題を解くことになり難しい問題となっている。

そこで、予めある合理的と思われる動作原理、仮説をたて、その仮説から演繹的に導かれる結果と実験結果を比べることでその仮説の妥当性を評価するという手法がしばしばとられる。このような立場の研究は、計算論的神経科学やサイバネティクスと呼ばれたりするが、本研究も計算論的神経科学の立場からピッチに関する心理学的知見を説明

[1] フレーム問題とは、状況に応じて適切な状態表現をどのように設定してやればよいかという問題である [1] [2]。

することを試みた研究である。

情報理論は、SN 比の向上やダイナミックレンジの増大、情報圧縮などの効率的な情報処理に役立つため、生物の各種感覚器官の動作原理としても合理的であろうと考えられ、古くからこの立場から情報表現を説明しようとする研究がなされてきた。Barlow は、感覚系ニューロンの目的は、入力情報の冗長性をできるだけ減らすような符号化を獲得することであると主張していた [7]。冗長性とは、実際のエントロピーを最大可能なエントロピーで割って 1 から引いた量であり、ある情報のエントロピーが最大可能なエントロピーからどれだけ離れているかを示す量である。冗長性が高いことは、エントロピーが取りえる最大のエントロピーと比較して小さいことを意味し、反対に冗長性が低いことは、エントロピーが大きいことを意味する。長い間、冗長性圧縮の仮説に基づくことで、一次視覚野に典型的に現われる線分抽出ニューロンの説明などはなされなかった。ところが、最近になって冗長性圧縮の原理と密接な関係を持つエントロピー最大化の原理 [8] [9][10] やスパースコーディングの仮説 [11] [12] に基づいた数理モデルを用いて、実際の視覚入力となる山や森などの自然画像を学習させると、線分抽出ニューロンの発現が説明できる、という研究が発表された。冗長性圧縮の原理自体は、感覚器官に依存しない原理であるので、聴覚のニューロンの受容野もこの原理に従って説明できることが期待される。本研究では、冗長性圧縮の原理に関連するエントロピー最大化、独立化の観点から聴覚一次野のニューロンの受容野 (基底) を学習することを試みた。

冗長性圧縮の原理は、心理学的知見を説明するのにも役立つ概念である。冗長性は、エントロピーで測られることからもわかるように情報の分布に偏りがあるかないかを測る指標になっている。情報の分布に偏りがある場合、予測的にその情報を推定することが可能である。このため、冗長性を圧縮した情報表現は、暗に冗長な情報に含まれていた予測可能な情報を利用することを示しており、予測的な推定と深い関係を持つ。自然界に存在する音は、冗長性を多く含んでいる。例えば、ヒトは普通、時間的に連続した音の系列を聞いている。音源は連続的に移動し、強弱の変化、高低の変化もたいてい連続的である。また、音は振動という物理現象によって生じるため、ある基本周波数で発生した音は同時にその整数倍の周波数の音も生じやすい。これらの自然音の物理的特性の存在は言い換えれば、我々が耳にする音が冗長性の高いものであることを示している。そして、我々はこれらの冗長性を利用することで、予測的な推定を行い高性能な音声認識を行っていると考えられる。錯覚現象のうちのいくつかは、普段、我々が耳にする音の性質を利用した予測的な推定を行っていると考えることで説明がつく。Virtual pitch (residue

pitch、または low pitch) は複数の周波数からなる複合音の音の高さ (ピッチ) が、その複合音には含まれていないある単一の周波数からなる純音と同じ高さをもつと知覚する錯覚現象である。本論文では、音の周波数に含まれている冗長性、すなわち周波数が互いにハーモニックな関係を持ちやすいという性質を利用した周波数情報の推定を行うことによって virtual pitch が生じることを示す。

2 章で、心理学的知見の説明と冗長性圧縮の概念が関係している他の例を示し、本研究で再現する現象である virtual pitch について詳しく説明する。また、聴覚系の解剖学的構造を概観し、心理学的知見と一致するような生理学的知見があることを説明する。

冗長性を圧縮するモデルとして、3 章でエントロピー最大化を行うモデルを提案する。このモデルでは、2 段階の階層的な処理によってエントロピー最大化が行われる。このモデルは、解剖学的な構造から考えて合理的な対応がとれるものであり、またその対応付けによって生理学的知見が説明できることを示す。

4 章では、3 章で提案した 2 段階の階層的な処理を非線形ノイズつき独立成分分析 (ICA) の定式化によって統一的に説明できることを示す。また、このモデルが線形独立成分分析やノイズつき独立成分分析を自然な形で包含していることを示す。このモデルによって、音高知覚の心理学的知見をより詳細に説明する。

5 章で、冗長性圧縮を行う情報表現を歪みあり符号の観点から説明する。入力系列をブロックごとに符号化するブロック符号は、どのブロック符号もベクトル量子化で表現が可能であり、さらにベクトル量子化は一般化された Lloyd アルゴリズム (GLA) などを用いることで局所最適性を保証する最適化が行える [13]。しかしベクトル量子化は符号化時に情報源の次元に対して指数的に計算量が大きくなる問題がある。ここでは、モデルに合理的な制約を加えた product code を採用することで、計算量を減らすことを考えた。この product code の構造を採用することで、歪みあり符号化の問題と冗長性圧縮の問題が密接な関係をもつようになる。product code の最適化法を提案し、得られる圧縮符号の性能を簡単なモデルで比較する。最後に 6 章で、本論文をまとめる。

第2章　聴覚の情報表現に関する心理・生理学的知見

本章では、音情報の統計的な性質と心理学的知見との密接な関係性について述べる。これによって概念的なモデルで説明されていた心理学的現象が、冗長性圧縮の原理に基づいた定量的な議論に焼きなおすことの妥当性を示す。また、聴覚の解剖学的構造を概観し、遠心性と求心性の両方の経路が豊富に存在していることを指摘する。後の章の数理モデルでは、この解剖学的構造との対応によるモデルの妥当性も議論する。

2.1. 心理学的知見

2.1.1 Bregman の発見的規則

2. 1 Bregman の発見的規則と冗長性

冗長性圧縮の概念によって聴覚系でみられる心理学的性質をいくつか説明することができる。Bregman は、聴覚系が解かなければいけない問題が不良設定問題となっていることから、生態学的に妥当性を有する拘束条件を発見的規則として用いていると考えた。Bregman は、以下の発見的規則を提案している [14] [15]。

Bregman の発見的規則

規則 1：変化は急激には起こらない。

・一つの音の属性は、滑らかにゆっくりと変化する傾向がある。

・同じ音源からの生じる一連の属性は、ゆっくりと変化する傾向がある。

規則 2：物が繰り返し振動する時には、共通の基本周波数の整数倍の音響成分が発生する。

規則 3：関連のない音が一緒に始まったり終わったりすることはない。

規則 4：一つの音響的事象に生ずる多くの変化は、その音を構成する各成分に同時に同じような影響を与える。

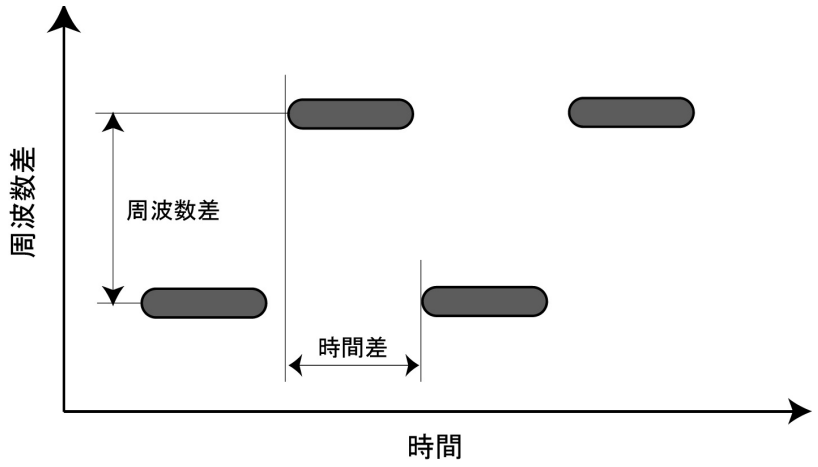


図 2.1 繰り返し交替する 2 音の提示

これらの規則は、それぞれ常に成立することが保証されているものではない。実際の環境で生じる多くの音がこれらの規則を満たしている、という統計的な規則に過ぎない。以下では、この Bregman の発見的規則と心理学的知見がどう関係するかについて述べる。

図 2-1 に示すように 2 つの音を代わる代わる提示する状況を考える。2 つの音の高さが異なる場合、提示する速度が遅い場合は、高い音と低い音がつながった一連の音の流れ（音脈）が聞こえる。しかし、2 つの音の提示間隔を短くしていくと次第に一つの音の流れとして知覚することが困難になり、継続する高い音に継続する低い音の 2 つの音脈が聞こえるようになる。図 2-2 は、この状況を音の繰り返し速度（一つの音の立ち上がりから次の音の立ち上がりまでの時間）と 2 音の音の高さの違いを表したものである。濃淡の濃い領域が、一つの音脈に聞こえた部分で、薄い領域が二つの音脈が聞こえた部分である。両者の中間の領域では、聞き方や前後の条件によって一つの音脈に聞こえたり、二つの音脈に聞こえたりする [16]。図のように 2 音の周波数差が大きく、また 2 音の交替する時間差が短いほど二つの別々の音脈として知覚されやすいことがわかる。このように音がいくつかの音脈に分離統合されることを音脈分凝と呼んでいる [17]。

音脈分凝が起こる理由としては、Bregman の発見的規則 1 の「一つの音の属性は、滑らかにゆっくりと変化する傾向がある」というのが挙げられる。自然音はこのような性質を持つので、脳はそれを利用して音脈の判断をしているものと考えられる。

規則 1 についてもう一つ例を挙げる。図 2-3 は、周波数変調音の知覚がその妨害音に

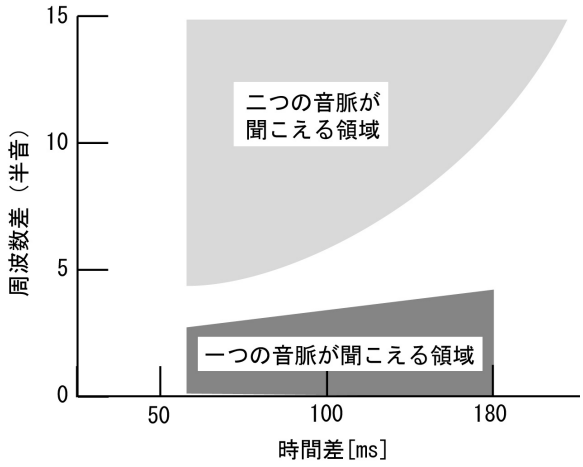


図 2.2 1つの音脈として聞こえる領域と2つの音脈として聞こえる領域。横軸は交替する時間差を表し、縦軸は2音の周波数差を表す。

よる影響をどのように受けるかを調べたものである。周波数変調音として、時間とともに次第に音が高くなっていくような音を与える。そして、一時的にそれに妨害音として帯域雑音を与える。もし、帯域雑音が十分強ければ、知覚される音は図2-3の下に示すように帯域雑音と周波数変調音の2つに分かれて聞こえる。もし、この部分の帯域雑音を取り除き、無音区間にすると周波数変調音が2回にわけて聞こえる。このことは、雑音の存在が一つの周波数変調音として知覚されるのに必要であったことを示している。これもやはり、規則1のような周波数の変化が継続しやすい傾向を利用し、このまま周波数の変化は続くであろうと予測することで帯域雑音と周波数変調音の2つにわけて知覚すると考えられる。

規則1では、同じ音源の音の変化はゆっくりとした連続的な変化でありやすいという時間的な性質を述べているのに対し、規則2の「物が繰り返し振動する時には、共通の基本周波数の整数倍の音響成分が発生する」という規則は、同じ音源の周波数間に存在する冗長性について述べている。たとえば、楽器が演奏する音も純音ではない。実際に楽器の音をスペクトル分解すると、複数の線スペクトルが存在していることがわかる。図2.4は、ピアノの C_3（いわゆる中央のドから1オクターブ低いドの音）を弾いた時のパワースペクトルを表示したものである[18]。

図のように広い周波数帯に強度が分散していることがわかる。その強度は、基本周波

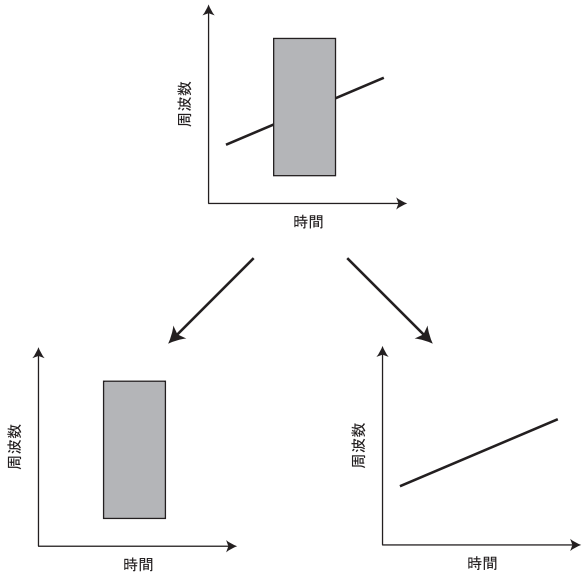


図 2.3 周波数変調音を帯域雑音によって妨害したときの知覚

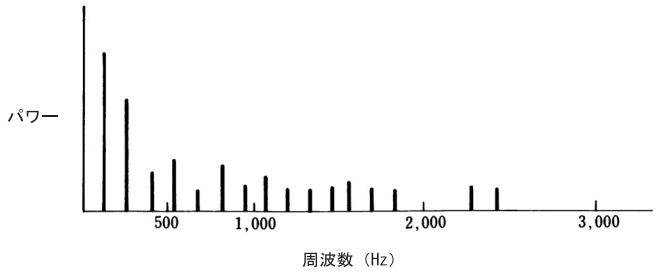


図 2.4 ピアノの C_3 のスペクトル (Lindsay & Norman, 1977 より引用)

数となる低周波数（この例では 131Hz）で最も強く、基本周波数の整数倍となる倍音の成分も強くなっている。全体としては、低周波数から高周波数になるにつれてスペクトル強度は弱くなっている。しかし、我々はこのようにいくつもの線スペクトルが存在する音をある高さ、この場合 131Hz の純音と同じ高さ、の一つの音として知覚する。この知覚による音の高さを物理的な音の高さ（周波数）と区別して、ピッチとよんでいる。ピッチは物理的には周波数とよく相関するが、周波数と一致しない場合もある。例えば、基本周波数を除いた倍音だけによる複合音を聞いた場合に、存在しない基本周波数の高さの音を知覚する。このようなピッチ感覚を得るのも規則 2 にあるような自然音の性質を利用した予測を行っているからだと考えられる。自然音は、基本周波数と共にその倍音

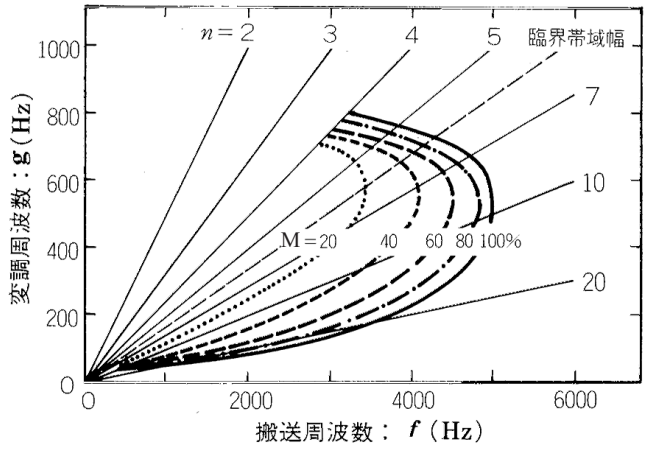


図 2.5 ピッチの存在する領域。横軸は高調波複合音のうちの中心周波数 f を表し、縦軸は、基本周波数 g を表す。M は変調度である。Ristma (1962) を基に引用。

を伴って現れるので、その性質を利用してハーモニックな成分からなる複合音を一つの音として認識を行っていると考えられる。カクテルパーティのような多くの人が話し合っている状況では、それだけ多くのスペクトルが混ざってしまう。しかし、それでも人は特定の人と会話を続けることができる。このような優れた音声認識にも、音声のハーモニックな構造を予測した音脈分凝が役に立っていると考えられる。実際、2 音を聞き分ける課題では、純音を聞き分けるより、それらの倍音を伴ったハーモニックな音を聞き分けるほうが成績のよいことが報告されている [19]。

2.1.2 ピッチに関する心理学的知見

ピッチは、音の周波数と密接に関係しているが、単純にピッチと周波数の対応がつくわけではない。前節でも述べたように、ほとんどエネルギーのない周波数に対してもピッチを知覚することがある [20][21]。ある単一の周波数成分をもつ音を純音と呼ぶが、その純音の周波数成分を含まないにも関わらず、純音と同じピッチを持つと知覚する現象を、virtual pitch(residue pitch、または low pitch) と呼ぶ。Virtual pitch は、基本周波数自体は含まないが、その倍音成分を含んでいるときに生じる。Virtual pitch の存在する領域は、Ristma によって調べられている [22]。

図 2.5 は、ピッチの存在する領域を表している。高調波複合音 $s(t)$ は、次のように 3 つの周波数成分からなる。

$$
\begin{aligned}
s(t) &= (1 + m\cos 2\pi g t)\sin 2\pi f t, \\
&= \frac{1}{2}m\sin 2\pi(f-g)t + \sin 2\pi f t + \frac{1}{2}m\sin 2\pi(f+g)t \qquad (2.1)
\end{aligned}
$$

t は時間、m は変調の度合いを表す。

$m = 0$ のとき、$s(t)$ は周波数 f のみからなる純音となり、$m = 2$ のとき、$s(t)$ は周波数 $f-g$, f, $f+g$ の 3 つの倍音からなる複合音となる。このとき、対応する基本周波数は g である。図 2.5 の変調度 M は、$M = \frac{m}{2}100 = 50m$ で与えられる。図からわかるように、

1. 倍音成分が弱いとき (変調度が小さいとき)、virtual pitch の知覚する領域は小さく、倍音成分が強いとき (変調度が大きいとき)、virtual pitch の知覚する領域は大きい。

2. 基本周波数 g が 100Hz 以下、800Hz 以上の領域では、virtual pitch はほとんど知覚されない。中間の 400Hz から 600Hz にかけての領域が最もよく virtual pitch が知覚される。

3. 中心周波数 f が大きくなると virtual pitch は知覚されない。

といった性質がある。

基本周波数は、virtual pitch を引き起こす複合音の周波数の差分、うなりに相当する周波数となっていることから、ハーモニックな複合音を聞いたときに蝸牛の基底膜上にそのうなりの成分が生じるのではないか、という説 (場所説) があった [23]。しかし、場所説に対しては以下の否定的な実験がある。

一つは、帯域雑音によって基本周波数の帯域をマスクする実験である。場所説の説明に従えば、基本周波数に対応する蝸牛の基底膜振動を妨害すれば virtual pitch は生成されないはずである。しかし、図 2.6B のように倍音の複合音に十分な強度の低周波数帯の雑音を加えたときにもピッチ感覚が生まれることが示された [24] [25]。

雑音の強度は、基本周波数の純音をマスクできる強度である。また、図 2.6C のように、倍音の成分の周波数帯域に同じ強度の雑音を加えると消滅する。

もう一つは、ハーモニックな複合音の周波数成分をシフトする実験である。場所説では、倍音の周波数の差分が基本周波数であることに依存して、virtual pitch が生成されるとしているので、倍音の周波数成分を全体的に上げ下げしても差分が変わらないので

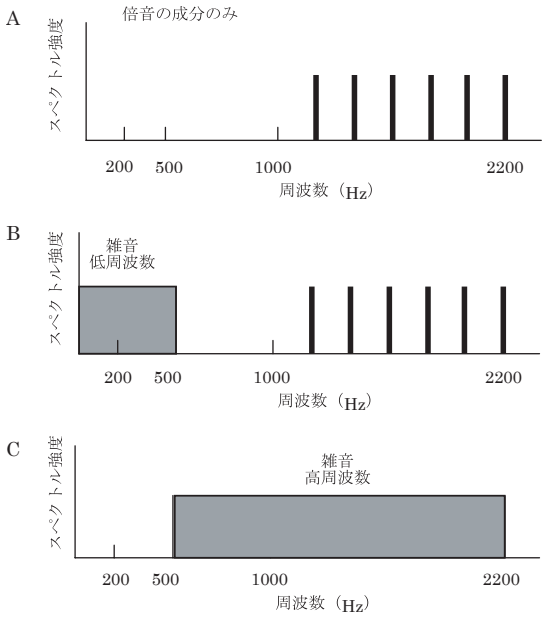


図 2.6　基本周波数を欠いた複合音の知覚。A:1 kHz,1.2 kHz,1.4 kHz,1.6 kHz,1.8 kHz,2.0kHz,2.2 kHz の音を聞かせたとき、200Hz の音の高さを感じる。B:低周波数に帯域雑音を与えても 200Hz は聞こえたように感じる。C:高周波数に帯域雑音を与えると 200Hz は聞こえなくなる。

ピッチも変わらないはずである。しかし、倍音成分をシフトするにしたがってピッチもシフトすること (ピッチシフト) が観測されている [26] [27]。

場所説と共に音高知覚理論で早くから提案されていた理論に時間説がある。時間説では、ピッチ感覚がいくつかのニューロンの発火のタイミングによって表現されているとしている。しかし、この発火のタイミングが複合音の周波数の位相関係に依存するのに対し、位相を変えてもピッチ感覚は損なわれないことが示されている [28][29]。さらに、ある基本周波数の整数倍の複合音を提示する代わりに片方の耳には、偶数次の複合音、もう一方の耳には、奇数次の複合音を提示したときにもやはりピッチ感覚が得られることからそもそも末梢の聴覚機構で説明することが困難であることが示された [30]。中枢に情報が伝えられるに従って、ニューロンの発火タイミングは不正確になるため、時間説による説明が困難であることも示唆している。次節では、場所説、時間説に代わる理論として提案されたパターン変換のモデルを説明する。

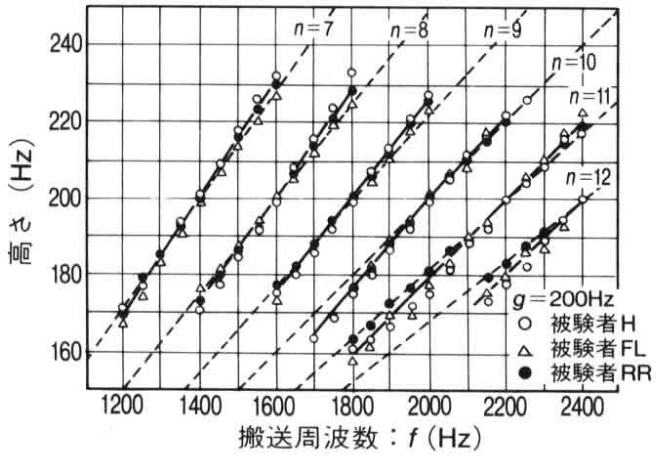


図 2.7 調波複合音の周波数をシフトしたときのピッチ。横軸は複合音の中心周波数を示し、縦軸はその複合音のピッチを示す。直線は実験値の直線回帰で、点線は第一次ピッチシフトと呼ばれる理論値。Schouten, et. al. (1962) を基に引用。

2.1.3 パターン変換モデル

パターン変換モデルの代表的なものは、Wightman [31]、Goldstein [32]、それに Terhardt[33] の 3 つのモデルがある。いずれのモデルもピッチの生成は、周波数分解された音信号が中枢のパターン処理でピッチを表す内部状態に変換されると説明している。図 2.8 は Wightman のモデルのブロックダイアグラムである。

まず、Wightman のパターン変換モデルについて説明する。入力音波形を $x(t)$、$x(t)$ を周波数解析したスペクトルパターンを $y(\omega)$ とする。周波数解析は、フーリエ変換した上で蝸牛の定 Q 解析 (constant Q analysis) を模擬するためそれぞれの周波数成分を対数周波数軸上で等しい大きさの三角フィルタで帯域フィルタリングしたものとしている。

ピッチパターン $s(\tau)$ は、このスペクトルパターン $y(\omega)$ にさらにフーリエ変換を施して得られる。すなわち、

$$s(\tau) = \frac{1}{2\pi} \int_0^\infty y(\omega) \cos \omega\tau d\omega \tag{2.2}$$

で得られる。

ピッチパターン $s(\tau)$ のうち、最大の値をとる $s(\tau_0)$ が抽出され、τ_0 の逆数がピッチの

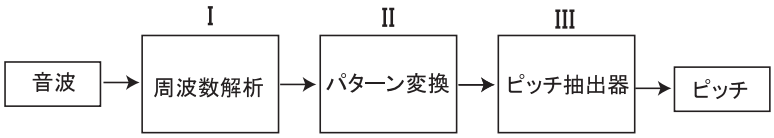


図 2.8 Wightman のパターン変換モデル。

周波数となる。このモデルでは、周波数のエネルギー成分を表すスペクトルパターンを変換してピッチパターンを得るので、エネルギーに影響を与えない音波の位相関係はピッチに影響を与えない。

次に、Goldstein の最適処理理論を説明する。Wightman のモデルと同じく、入力音波形 $\mathbf{x}(t)$ はスペクトルパターン $\mathbf{y} = [y_1, y_2, \cdots, y_{\mathrm{n}}]'$ に変換されて中枢の入力となる。スペクトルパターンは Wightman のモデルとは違い、n 個のスペクトルに離散化されていることをベクトルで表している。ピッチを s としたときの尤度 $p(\mathbf{y}|s)$ が次式で与えられる。

$$\begin{aligned} p(\mathbf{y}|s,j) &= \prod_{i=1}^{n} p(y_i|s), \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(y_i-(j+i-1)s)^2}{2\sigma_i^2}\right) \end{aligned} \tag{2.3}$$

ただし、σ_i はある定数で、j は入力スペクトルが基本周波数の何倍で表現できるかを決める自然数である。尤度 $p(\mathbf{y}|s)$ を最大にする (s, j) の組が求められ、そのときの s がピッチとして与えられる。

最後に、Terhadt の virtual pitch 理論を説明する。まず、入力音波形 $x(t)$ は、離散的な

周波数成分			約数の値
800	1000	1200	
400	500	600	2
266.7	333.3	400	3
200	250	300	4
160	***200***	240	5
133.3	166.7	***200***	6

表 2.1 Terhadt の virtual pitch 理論の説明

スペクトルパターン $\mathbf{y} = [y_1, y_2, \cdots, y_{\mathrm{n}}]'$ に変換される。さらに変換されたスペクトルパターンは、それぞれスペクトルの強度を表すものではなくそのスペクトルが含まれているかいないかの 2 値情報となっているとする。例として、周波数成分に 800,1000,1200Hz が含まれていたとされた場合、どのようにピッチを推定するか表 2.1 を用いて説明する。まずそれぞれの周波数成分を自然数で割った値を求める。そして、その商として得られた値のうち最も一致する回数が多いものがピッチとして推定される。この例の場合、200 が最も一致する回数が多い値 (3 回) なので 200Hz がピッチとして推定される。

これら 3 つのモデルは、数学的に密接的に関係していることが示されている [34][35]。スペクトルパターンは、連続値の場合、離散値の場合どちらの場合でも説明できるが、観測に限界がある生物学的な制約から考えると離散値の方がより自然な仮定であるので、スペクトルパターンは離散値で与えられるとする。Terhadt のモデルは、Goldstein のモデルにおいて二乗誤差 $\frac{(y_i-(j+i-1)s)^2}{2\sigma_i^2}$ で評価されていた部分を $\Gamma((y_i-(j+i-1)s)^2)$ のように 2 値関数 Γ を用いた表現に置き換えたとき、Goldstein のモデルと一致する。また、Goldstein のモデルにおいて自然数 j を 1 から最大で m の間で尤度を最大にする解 (s, j) を探索するとする。このとき、スペクトルパターン $\mathbf{y}$ の最低周波数を f_0 とすると、基本周波数は $\frac{1}{j}f_0$ $(j = 1, \cdots, m)$ の m 通りの値から探索していることになる。そこで、s_j を周波数 $\frac{1}{j}f_0$ のピッチの強さを表す変数とし、ベクトル $\mathbf{s}$ を $\mathbf{s} = [s_1, \cdots, s_{\mathrm{m}}]'$ とする。列ベクトル $a_j = [j, \cdots, n+j-1]'$ を定義すると、Goldstein のモデルは、$\mathbf{y} - \mathbf{a}_j s_j$ の重み付き二乗誤差 $\sum_{i=1}^{n} \frac{1}{2\sigma_i^2} (y_i - \mathbf{a}_j s_j)^2$ を最小にする変数 s_j を求め、その s_j に対応する周波数をピッチとしていることがわかる。Terhadt のモデルの式 (2.2) は、スペクトルパターンとピッチパターンをそれぞれ離散化して $\mathbf{y}$ と $\mathbf{s}$ で表すと、$\mathbf{s} = W\mathbf{y}$ の線形変換で表現できる。この線形変換によって得られる $\mathbf{s}$ のうち最大のものを選び、その $\mathbf{s}$ に対応する周波数をピッチとしている。Terhadt のモデルの場合、この線形変換は正方行列で逆行列をもつので W の逆行列を A とすると、二乗誤差 $(\mathbf{y} - A\mathbf{s})^2$ を最小にする $\mathbf{s}$ を求めることと等価である。従って Goldstein のモデルでは、ある一つの高調波複合音でスペクトルパターンを説明しようとすることと、尤度を最大、すなわち重み付き二乗誤差を最小にする s_j に対応する周波数をピッチとしているのに対し、Terhadt のモデルでは、複数の高調波複合音でスペクトルパターンを説明しようとしていることと、二乗誤差を最小にする $\mathbf{s}$ を求めたときにもっと大きい値をとる s_j に対応する周波数をピッチとしている点で異なる。しかし、いずれも離散化したとき、内部変数 $\mathbf{s}$ に線形変換をかけてスペクトルパターンを説明しようとしている点では基本的に近い関係にある。

音声分析の分野では、古くからケプストラム分析が用いられてきた [36][37]。ケプストラムは分析は、Wightman のモデルにおいてスペクトルパターンをフーリエ変換後の振幅スペクトルの対数とすれば一致する。人の音高知覚を説明しようとする理論が、人の音声分析の手法と近いものであることは興味深い。

2.2. 生理学的知見

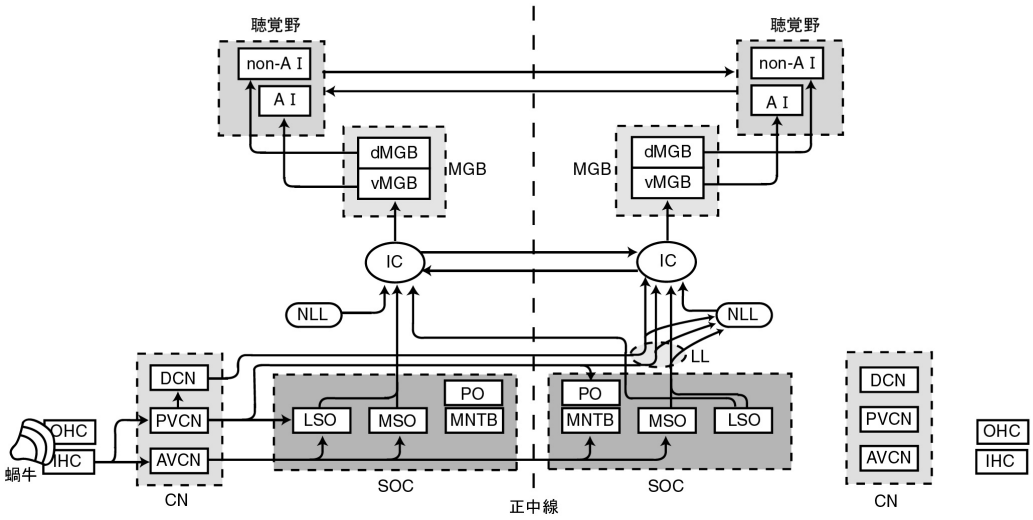
2.2.1 初期聴覚系の解剖学的知見

聴覚の伝導路は、視覚に比べ中継する神経核が多く複雑である。ここでは、本研究のモデルと対応を考えるため、蝸牛や聴神経線維について詳しく述べ、その他については簡単な説明に留める。聴覚において入力となる刺激は、音、つまり空気の疎密波である。聴覚は、この物理的な圧力変化である機械信号をまず、電気信号に変換する必要がある。この変換を行っているのが蝸牛という器官である。鼓膜から耳小骨連鎖を経た音の振動は、蝸牛の前庭窓を振動させる。この振動は、進行波となって蝸牛内のリンパ液を伝播し、それが基底膜上に進行波を引き起こす。この基底膜上の進行波は、基底膜の物理的特性の異方性により、蝸牛の入口（基部）から出口（頂部）にかけてその振幅が変化する。その変化は、基底膜上のある場所に近づくにつれて振幅が大きくなり、そこを過ぎると急激に小さくなるというものである。その基底膜上で最大振幅をとる場所というのが、入力音の周波数に依存しており、入力音の周波数情報は、基底膜上の最大振幅をとる位置として場所表示されることになる。そして、この基底膜上に存在する有毛細胞が基底膜の振動に応じてイオンチャネルを開き、有毛細胞の脱分極（定常状態では細胞外に比べて低い細胞内電位が高くなること）を起こす。有毛細胞の脱分極は、神経伝達物質の放出を促し、聴神経線維の内部電位を上昇させる。このようにして蝸牛において音の機械的信号が周波数分解された後に電気信号として中枢に伝達される。しかし、実は蝸牛の働きは単なる受動的な信号の変換装置ではない。蝸牛の有毛細胞には、内有毛細胞と外有毛細胞の 2 種類が存在している。ヒトでは内有毛細胞が約 4000 個、外有毛細胞が 3 倍の約 12000 個ある。この数の多い外有毛細胞には細胞自体が伸縮するという運動機能がある（内有毛細胞にはこのような運動機能はない）。この外有毛細胞の能動的な働きによって、蝸牛は鋭敏な周波数同調を可能にすると考えられている。内有毛細胞は外有毛細胞に比べて数は少ないが、蝸牛から中枢へ向かう求心性のニューロンのほとんどが内有毛細胞とシナプスを作っている。聴神経線維は、I 型聴神経線維 II 型聴神経線維の 2 種

類ある。I 型聴神経線維はヒトでは約 32000 本（大人）から約 41000 本（子供）あり、ミエリン鞘に包まれ（有髄）電気的に絶縁されている。II 型聴神経線維は、ヒトで約 3500 本あり、無髄で電気的に絶縁されていない。この数の多い I 型聴神経線維が求心性で主に内有毛細胞と接続しており、数の少ない II 型聴神経線維が遠心性で主に外有毛細胞と接続している。また I 型聴神経線維は分岐せず 1 個の内有毛細胞に接続し、1 個の内有毛細胞に対し平均して約 10 本の聴神経線維が接続しているが、II 型聴神経線維は分岐して約 10 個の外有毛細胞に接続し、1 個の外有毛細胞は 5、6 本の聴神経線維と接続している。したがって内有毛細胞からは多くの神経線維が発散し、外有毛細胞には発散と収束の両方がみられる。有毛細胞はまた、オリーブ蝸牛束からの遠心性の投射も受けている。I 型聴神経線維は聴神経ともよばれその反応が調べられているが、II 型聴神経線維についての研究はほとんどなく、その性質はよくわかっていない。聴神経の発火は、それがシナプスしている内有毛細胞の脱分極に依存する。聴神経の発火が増加する最低の入力音の音圧を周波数に対して描いた図を周波数同調曲線というが、聴神経の同調曲線は V 字型をしており、ある周波数に対してよく応答する特性をもつ。最も低い音圧で応答する周波数をそのニューロンの特徴周波数と呼んでいる。また、聴神経は、1kHz 程度までの周波数の入力音に対してはその位相に同期して発火することができ、位相同期発火と呼んでいる。聴神経線維からの投射は、蝸牛神経核に及んでいる。蝸牛神経核のニューロンの特徴周波数を調べると、特徴周波数の低いものから高いものへと順番にニューロンが並んでおり、このような規則的配列をトノトピシティと呼んでいる。このトノトピシティは、大脳皮質の一次聴覚野まで保存されている。蝸牛神経核のうち、前方蝸牛腹側核のニューロンは聴神経とよく似た性質をもつニューロンが多いの対し、背側蝸牛神経核では様々な応答パターンをもつニューロンがみられ、複雑な信号分析を行っているのではないかと考えられている。ニューロンの同調曲線も V 字型の単峰性のものだけでなく、抑制によって興奮性の範囲が島状に分離しているものもある。刺激音に対する選択性も特徴周波数の純音だけでなく、周波数変調音に強く応答するニューロンも見つかっている。背蝸牛神経核からは上オリーブ核と下丘へ投射する経路にわかれる。上オリーブ核は両耳からの神経線維が初めて統合され、音源定位に関与すると考えられている。そのニューロンの発火も両耳間の刺激音の到達時間差、音圧差に選択的に反応するものが見つかっている。下丘は、内側膝状体や上オリーブ核、蝸牛神経核、さらには体性感覚系からの求心性、遠心性の線維が集まっている。下丘の下丘中心核でもトノトピシティがみられ、蝸牛神経核と同程度または、それ以上に鋭い周波数同調曲線を見せるニューロン

もある。しかし、位相同期発火特性を示すニューロンは激減し、同調できる周波数は最高でも 600Hz 程度である。また、ニューロンの発火も周波数変調音や振幅変調音に選択的な応答を示したり、両耳間の時間差、音圧差に選択性を持ったりと様々である。内側膝状体は、視床にある聴覚系の中継核で、大脳皮質と下丘から遠心性、求心性の投射をうける。内側膝状体腹側核でもトノトピシティがみられ、同調曲線も鋭く、下丘でみられた極端に鋭い同調曲線をもつニューロンは見られないが、聴神経と同等またはそれ以上の同調曲線をもつニューロンが見つかっている。内側膝状体にまで高次になると、ニューロンの発火パターンは複雑で入力音の性質から分類することはますます難しくなってくる。無麻酔で自由に動き回る動物からの記録では、ニューロンの応答自体が不安定で時間によって興奮と抑制の帯が時間によって現れたり消滅したりするという報告もある [38]。聴覚野は、内側膝状体から求心性の投射をうける大脳皮質の領域である。一次聴覚野は、内側膝状体からの投射をうけており、トノトピシティが存在する。聴覚野のニューロンは複雑な反応を示すものが多い。なお、聴覚一次野のニューロンの最適周波数は、短時間の学習で変化したり [39][40]、末梢系の損傷によってトノトピシティの周波数地図が変化したりするという可塑性も観察されている [41]。このように、聴覚伝導路はいくつもの中継核を経由して大脳皮質に至る。全体としてみると、その経路は求心性のフィードフォワードの結合だけでなく、遠心性のフィードバックの結合も同じくらい豊富に存在しており、その遠心性の支配は音刺激を電気信号に変換する蝸牛にまでさかのぼることがわかる。そのため、フィードバックの経路も重要な働きを担っていることが考えられる。図 2.2.1 に主要な聴覚伝導路を模式図で示す。

A. 求心性



B. 遠心性

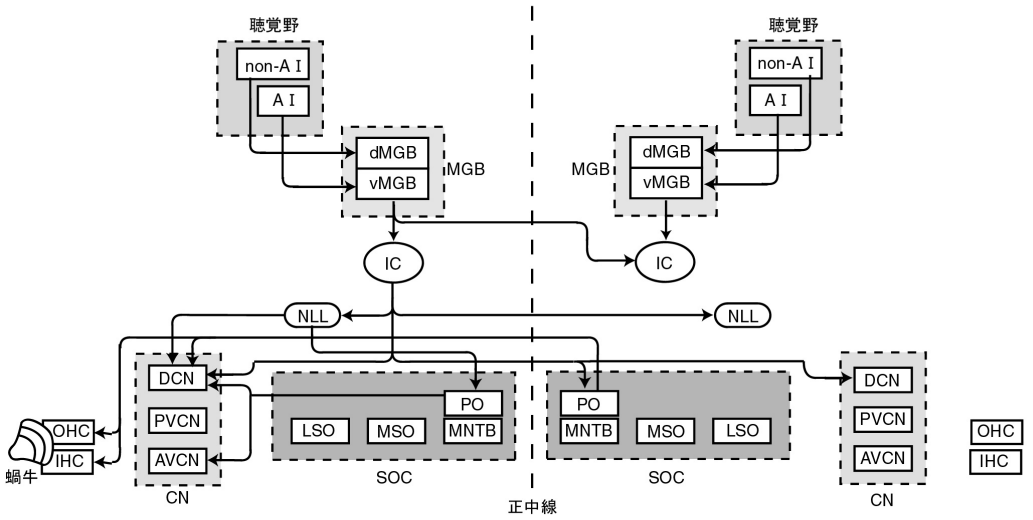


図2-8　主要な聴覚伝導路の模式図　A：求心性経路　B:遠心性経路
IHC：内有毛細胞，OHC：外有毛細胞，CN：蝸牛神経核，AVCN：前方腹側蝸牛神経核，PVCN：後方腹側蝸牛神経核，DCN：背側蝸牛神経核，SOC：上オリーブ複合体，LSO：上オリーブ外側核，MSO：上オリーブ内側核，MNTB：台形体内側核，PO：周辺核，LL：外側毛帯，NLL：外側毛帯核，IC：下丘，MGB：内側膝状体，MGBv：内側膝状体腹側核，MGBd：内側膝状体背側核，
A I ：一次聴覚野，non-A I ：一次聴覚野以外の聴覚野

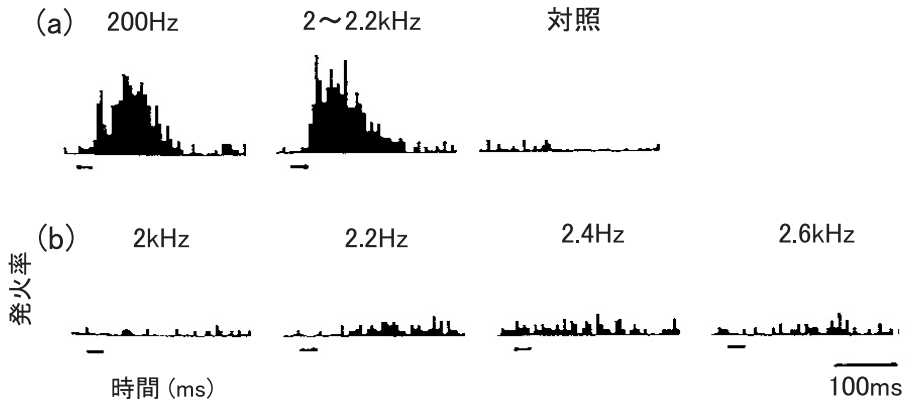


図 2.9 サルの聴覚一次野のニューロンの純音と複合音に対する応答。2000Hz、2200Hz、2400Hz、2600Hz の音を同時に聞かせたとき、サルの聴覚一次野のあるニューロンは、基本周波数の 200Hz の純音を聞かせたときと同様に応答する。しかしながら、その高調波複合音のうちの 1 つの周波数成分だけを聞かせてもニューロンの応答は小さい。Riquimaroux & Hashikawa (1994) を基に引用。

2.2.2 ピッチに関する聴覚一次野のニューロンの応答特性

一次聴覚野のニューロンの特徴周波数が連続的に並ぶトノトピシティがあることを前節で述べた。ここでは、純音ではなくピッチを引き起こす高調波複合音を提示したときのニューロンの活動に関する知見を述べる。Pantev[42] は、脳磁図 (MEG:magnetoencephalography) を用いた実験によりヒトの聴覚一次野のある部分は、純音および virtual pitch をもたらす高調波複合音のどちらにも応答するが、高調波複合音のうちのある高周波数音を一つだけ取り出して聞かせても応答しないことを報告している。また、Riquimaroux らはこれをより時空間分解能の高い電気生理実験で確かめている [43]。図 2.9 に示すように、Riquimaroux らはサルの一次聴覚野ニューロンに基本周波数のみの純音を聞かせた場合と、その倍音を 4 つ含む高調波複合音を聞かせた場合に同じように応答するニューロンを観察している。また、高調波複合音を構成する高周波数を一つ取り出して聞かせてもほとんど応答しないことがわかる。

また一次聴覚野近辺に障害をもつピッチ感覚は損なわれることが知られている [44]。これらの知見から、聴覚一次野のニューロンによってピッチが表現されていると考えるのは、少なくとも実験結果と矛盾を起こさない妥当な仮説であると考えられる。

第3章　エントロピー最大化による音高知覚モデル

ここでは、エントロピー最大化を2段階のステップに分離して階層的にエントロピー最大化を行うモデルを提案する。このモデルで virtual pitch のシミュレーションが可能であると共にモデルと解剖学的構造との対応から聴神経でみられるマスキングの定性的説明が可能であることを示す。

3.1. 数理モデル

3.1.1 モデル構造

図 3-1 に示されるモデルは、前処理を含めると、3つの階層からなる。第1階層において、入力された音は高速フーリエ変換 (FFT) によって周波数情報に変換される。この周波数変換は蝸牛の機械的な処理に対応すると仮定している。第2階層では、第1階層で得られた周波数情報が周波数ごとに非線形に変換される。これは、内有毛細胞において機械的信号を電気的信号に変換する処理に対応すると仮定している。第3階層において、第2階層で電気的信号として表現された周波数情報が脳内の内部表現に変換される。この第3階層は、さらに二つの部分構造 (レイヤー) からなる。上位レイヤーからのフィードバック信号は、トップダウンの予測を表現し、下位レイヤーからのフィードフォワード信号は、そのフィードバックによる予測と第2階層からの入力との誤差を表現する。フィードフォワード経路とフィードバック経路は、それぞれ聴神経とオーリブ蝸牛束などに対応すると仮定している。上位レイヤーにおける表現が、全体のモデルの出力であり、これが中枢の聴覚システムに与えられる脳内内部表現となっている。ここでは、周波数情報の処理に注目しており、時間的な変化に伴う影響や両耳性の影響などは無視している。我々のモデルは、2.2 節で述べた聴覚伝導路の解剖学的知見に矛盾するものではなく、音の周波数情報処理の重要な部分を抜き出したモデルと考えることができる。

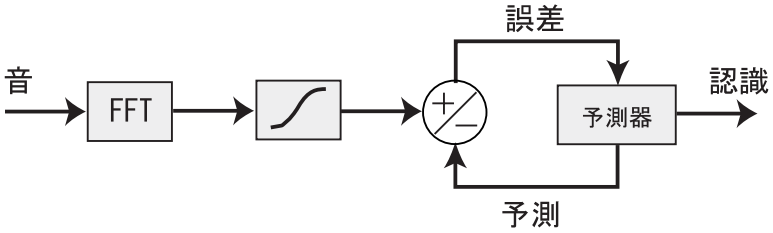


図 3.1 エントロピー最大化のためのモデル構造。第 1 階層において、入力音は FFT によって周波数分析がなされる (蝸牛に対応)。第 2 階層では、周波数情報は電気的信号に非線形に変換される (内有毛細胞に対応)。第 3 階層では、電気的信号は中枢の聴覚システムのために内部表現に変換される。第 3 階層は、上位および下位の 2 つのレイヤーからなる。上位レイヤーからのフィードバック信号 (オリーブ蝸牛束に対応) は、トップダウンの予測を表し、下位レイヤーからのフィードフォワード信号 (聴神経に対応) は、トップダウン予測と第 2 階層からの入力との誤差を表す。

3.1.2 情報理論による解釈

情報理論によれば、n 次元 確率変数 $\mathbf{x} = [x_1, x_2, \cdots, x_{\mathrm{n}}]'$ の観測によって得られる平均情報量は $\mathbf{x}$ の結合エントロピー $H(\mathbf{x})$ で与えられる。

$$H(\mathbf{x}) = -\int p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x} \tag{3.1}$$

ここで、$p(\mathbf{x})$ は $\mathbf{x}$ の確率密度関数を表す。プライム記号 ($'$) は転置を表す。結合エントロピーは次のように分解できる。

$$\begin{aligned} H(\mathbf{x}) &= \sum_{i=1}^{n} H(x_i) - \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{\tilde{p}(\mathbf{x})} d\mathbf{x} \\ &= \sum_{i=1}^{n} H(x_i) - D[p(\mathbf{x}) \parallel \tilde{p}(\mathbf{x})] \end{aligned} \tag{3.2}$$

ここで、$H(x_i)$ は x_i の周辺エントロピー $H(x_i) = -\int p(x_i) \log p(x_i) dx_i$、 そして、$\tilde{p}(\mathbf{x})$ は要素ごとの周辺分布の積 $\tilde{p}(\mathbf{x}) = \prod_{i=1}^{n} p(x_i)$、$p(x_i) = \int p(\mathbf{x}) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$ である。また、 $D[p \parallel q]$ は確率密度分布 p と q の相対エントロピー (Kullback-Leibler ダイバージェンス) を表す。式 (3.2) のように、結合エントロピーは二つの項の差として表現できる。第一項は周辺エントロピーの和であり、第二項は要素間の独立性からのずれの度合いを表す。われわれのモデルでは、結合エントロピーを最大化することを目的として、式 (3.2) の 第一項の最大化と、第二項の最小化を階層的に行う。これら二つのプ

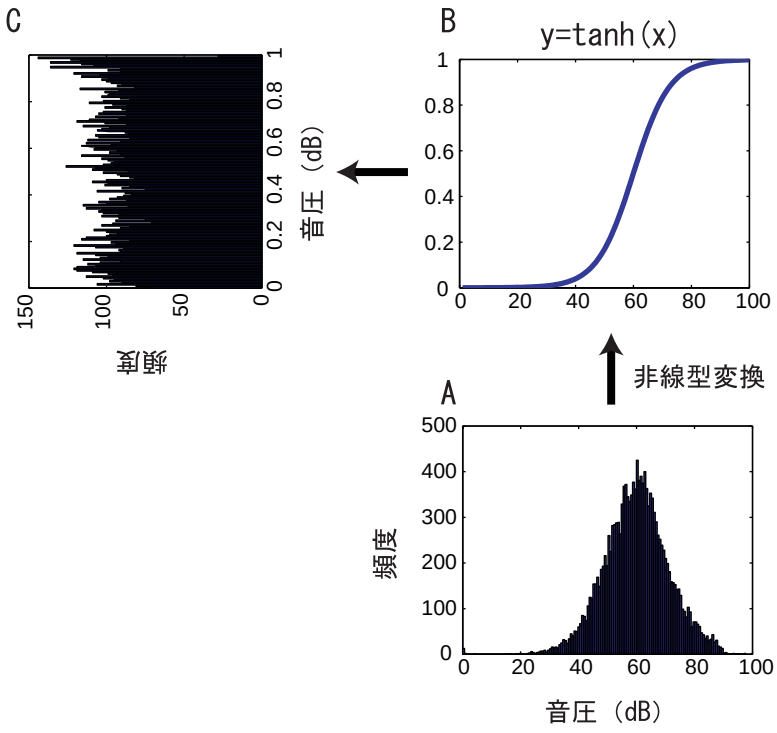


図 3.2 周辺エントロピー最大化のための非線形変換。A：変換前の入力の頻度分布。入力は 1034Hz の対数周波数強度。B：変換に用いた非線形関数 y=tanh(x)。C：変換後の頻度分布。

ロセスは、工学的にはそれぞれ非線形変換とノイズあり ICA により実現され、以下で詳しく述べる。

3.1.3 周辺エントロピー最大化

まず、第 2 階層で行われる周辺エントロピーの最大化について議論する。スカラー変数 y は、スカラー変数 x を連続で可逆な関数 f を用いて $y = f(x)$ のように変換して得られるものとする。ここで、 x は FFT によって得られたある周波数成分のパワーの対数であるものとする。このとき $H(y)$ と $H(x)$ は以下の関係をもつ。

$$H(y) = H(x) + \int p(x) \log |\frac{d}{dx} f(x)| dx \tag{3.3}$$

音楽などの自然音の場合、ある周波数成分のパワーの対数についての分布は図 3.2A に示すように正規分布のような単峰な分布となりやすい [45]。入力 x が正規分布に従うな

らば、エントロピー $H(y)$ を最大化する非線形変換は、正規分布の累積密度関数である誤差関数となる。シグモイド関数は誤差関数と同じような形になるため、図 3.2B に示すようにそれを非線形関数 $f(x)$ に選ぶことで、おおむね正規分布に従う対数パワーの非線形変換を行った。特に値域が $[0, 1]$ となるように、$y = f(x|\mathbf{w}) = (\tanh(w_1 x + w_0) + 1)/2$ と定義した。ここで、定義域を無制限にするとエントロピーの値は任意にできるので、エントロピーの大小比較のためには定義域に制限を加える必要があることに注意する。

$f(x|\mathbf{w})$ で用いられるパラメータ $\mathbf{w}$ は、エントロピー最大化基準から決定される。入力のエントロピー $H(x)$ はパラメータ $\mathbf{w}$ に依存しないことに注意して、エントロピー $H(y)$ の最大化を勾配法 $\triangle w \propto \partial H(y)/\partial w$ によって行うと、

$$\triangle w_1 \propto \frac{1}{w_1} - 2\langle x(2y-1)\rangle \approx \frac{1}{w_1} - 2x(2y-1) \tag{3.4}$$

$$\triangle w_0 \propto -2\langle y\rangle + 1 \approx \ -2y + 1 \tag{3.5}$$

となる。この学習則は、Bell と Sejnowski による infomax ICA [46] のスカラー版である。彼らは結合エントロピーの最大化に発展させて ICA の問題を解いているが、ここでの式 (3.4)、(3.5) はベクトル $\mathbf{x}$ のそれぞれの要素の周辺エントロピーを最大化するために用いている。図 3.2 からもわかるように、エントロピーを最大にする変換によって出現頻度の高い情報を予測し、そこに予め多くのリソース (発火率による表現) を割いていることがわかる。

図 3.3 に示すようにシグモイド状の非線形な応答関数は有毛細胞で観察されている [47] 他、他の感覚器官にも共通する応答である [48] ため、第 2 階層においてシグモイド関数を用いて行う非線形変換は生物学的にも妥当と考えられる。

3.1.4 ノイズ付き独立成分分析

第 2 階層において要素ごとの周辺エントロピー最大化により得られる変換信号 $\mathbf{y}$ は、第 3 階層に与えられる。第 3 階層は上位および下位のレイヤーからなるネットワークであり、$\mathbf{y}$ の要素間の依存性を削減した内部表現 $\mathbf{s}$ を生成する。この処理は、情報理論上での冗長性圧縮の原理 [7] に基づくだけでなく、2 章で述べた心理学的な仮説とも関係がある。先述のように Bregman [17] は、人間は別々の音源を適切に区別できるように音の体制化、音脈分凝を行っていると提唱した。音脈分凝のための内部表現は、別々の音源は独立に音を発生すると仮定できるならば、ブラインド音源分離、すなわち ICA によって

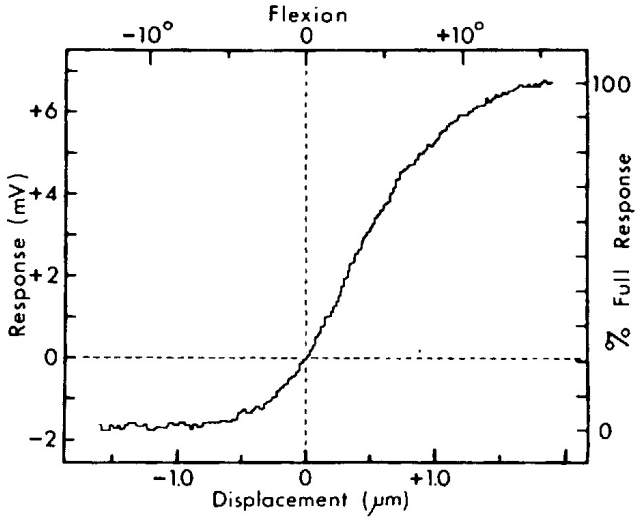


図 3.3　ウシガエル球形嚢の有毛細胞の不動毛変位と電位変化の関係。有毛細胞の応答は、不動毛変位に対して非線形な電位の飽和特性を示す。Hudspeth & Corey (1977) より引用。

得られる。この節では、事後確率最大化 (MAP：maximum *a posteriori*) 推定の定式化を用いて ICA の問題を解くことで、この依存性削減を実現する。

第 3 階層のネットワークは以下の確率的生成モデルを用いて記述される。

$$\mathbf{y} = A\mathbf{s} + \mathbf{n} \tag{3.6}$$

$\mathbf{s}$ は内部表現を表しこれがネットワークの出力となる。簡単のため、$\mathbf{s}$ の次元は $\mathbf{y}$ の次元である n と一致すると仮定する。$\mathbf{n}$ は、平均 0、分散 $\sigma^2 I_n$ の正規分布に従うノイズである。I_n は n 次元の単位行列を表し、パラメータ A は $n \times n$ 行列である。

$\mathbf{y}$ の系列 $Y = \{\mathbf{y}_1, \mathbf{y}_2 \cdots, \mathbf{y}_T\}$ を観測した後、パラメータ A の MAP 推定は次式の対数事後確率の最大化により得られる。

$$\begin{aligned} \log p(A|Y) &= \sum_{t=1}^{T} \log p(\mathbf{y}_t|A) + \log p(A) + const \\ &= \sum_{t=1}^{T} \log \int p(\mathbf{y}_t, \mathbf{s}_t|A) d\mathbf{s}_t + \log p(A) + const. \end{aligned} \tag{3.7}$$

パラメータ A の事前分布として、要素ごとの正規分布を仮定した。

$$p(A) = \prod_{i,j=1}^{n} p(A_{i,j}) = \prod_{i,j=1}^{n} \exp(-\frac{\lambda}{2} A_{i,j}^2) \tag{3.8}$$

ここで、 ハイパーパラメータ $\lambda(>0)$ はある定数である。式 (3.7) の対数事後確率には隠れ変数 $\mathbf{s}$ が含まれるので、式 (3.7) のパラメータ A についての最大化には、以下の EM アルゴリズム [49] を用いた。

- E ステップ

  現在のパラメータ推定値 $\bar{A}$ を用いると、期待対数尤度 (厳密には期待対数事後確率) は

$$Q(A|\bar{A}) \quad = \quad \frac{1}{T}\sum_{t=1}^{T}\int p(\mathbf{s}_t|\mathbf{y}_t,\bar{A})\log p(\mathbf{y}_t,\mathbf{s}_t|A)d\mathbf{s}_t + \log p(A) \tag{3.9}$$

  と書ける。ただし、$p(\mathbf{s}_t|\mathbf{y}_t,\bar{A})$ は以下の隠れ変数の事後分布である。

$$p(\mathbf{s}_t|\mathbf{y}_t,\bar{A}) = \frac{p(\mathbf{y}_t,\mathbf{s}_t|\bar{A})}{\int p(\mathbf{y}_t,\mathbf{s}_t|\bar{A})d\mathbf{s}_t} \tag{3.10}$$

- M ステップ

  期待対数尤度 $Q(A|\bar{A})$ をパラメータ A について最大化する。

式 (3.10) の隠れ変数の事後分布を解析的に求めるのは困難であり、また生物学的にも自然でないので、事後分布は隠れ変数の MAP 推定値 $\hat{\mathbf{s}}_t$ において鋭いピークを持つデルタ関数として近似できると仮定した。この近似を用いると、以下で示すように学習アルゴリズムが局所的なヘブ学習則として表され、生物学的にも自然なものとなる。この MAP 近似により、式 (3.9) の期待対数尤度は、

$$Q(A) = \frac{1}{T}\sum_{t=1}^{T}\log p(\mathbf{y}_t,\hat{\mathbf{s}}_t|A) + \log p(A) \tag{3.11}$$

となり、これを最大化する新しいパラメータ A は、

$$A = \langle \mathbf{y}\hat{\mathbf{s}}'\rangle_T(\langle \hat{\mathbf{s}}\hat{\mathbf{s}}'\rangle_T + \lambda\sigma^2 I_n)^{-1} \tag{3.12}$$

として解析的に求められる。ここで、$\langle f(\mathbf{x},\hat{\mathbf{s}})\rangle_T \equiv \frac{1}{T}\sum_{t=1}^{T} f(\mathbf{x}_t,\hat{\mathbf{s}}_t)$ である。

式 (3.12) は、観測系列 Y 全体に対する十分統計量を用いるバッチ学習則である。一方で、パラメータを観測データが得られるたびに更新するオンライン学習は、バッチ学習に比べ生物学的により自然な学習であると思われる。

オンライン EM アルゴリズム [50] によると、M ステップで必要になる十分統計量は以下のように逐次的に計算することができる。

$$\langle f(\mathbf{x},\hat{\mathbf{s}})\rangle_t \;=\; \langle f(\mathbf{x},\hat{\mathbf{s}})\rangle_{t-1} + \eta(t)(f(\mathbf{x}_t,\hat{\mathbf{s}}_t) - \langle f(\mathbf{x},\hat{\mathbf{s}})\rangle_{t-1}) \tag{3.13}$$

ここで、 $\eta(t)$ は確率近似法で用いられるのと同じく 0 に減衰していく学習係数である。共分散行列の逆行列 $(\langle\hat{\mathbf{s}}\hat{\mathbf{s}}'\rangle_t + \lambda\sigma^2 I_n)^{-1}$ を逐次的に更新するよう書き直して、$\eta(t)^2$ が掛けられている項を十分に小さいとして無視すると、M ステップにおけるパラメータ更新則は

$$\begin{aligned} A_t \;&=\; A_{t-1} + \eta(t)\{(\mathbf{y}_t - A_{t-1}\hat{\mathbf{s}}_t)\hat{\mathbf{s}}_t' - \lambda\sigma^2 I_n\} \\ &\quad\times(diag(\langle\hat{\mathbf{s}}\hat{\mathbf{s}}'\rangle_{t-1}) + \lambda\sigma^2 I_n)^{-1} \end{aligned} \tag{3.14}$$

となる。ただし、A_t は t 番目のデータ $\mathbf{y}_t$ を観測した後のパラメータ推定値で、$diag(M)$ は行列 M と対角成分が同じ対角行列である。この学習則は正規化付きのヘブ学習の形となっており、更新によって加えられる項は主に予測誤差 $(\mathbf{y}_t - A_{t-1}\hat{\mathbf{s}}_t)$ と内部表現の予測値 $\hat{\mathbf{s}}_t$ の内積に依存している[1]。

式 (3.14) の学習則は、視覚系における Olshausen と Field の研究 [11] や、Rao と Ballard の研究 [12] で用いられた学習則と似ている。しかしながら、式 (3.14) の学習則では従来の学習則に $\lambda\sigma^2 I_n$ という正規化の役割を果たす項が付加されている点が異なる。また式 (3.14) の学習則は、Belouchrani と Cardoso [51] が導出したノイズ付き ICA の学習則とも似ている。違いは、パラメータ A に対する事前分布が付加されているかどうかと、ヘブ則として表現できるようにするためにオンライン型の学習則への近似が行われているかどうかである。

一方で、E ステップで用いられる $\mathbf{s}$ の MAP 推定値は、

$$\begin{aligned} \hat{\mathbf{s}}_t \;&=\; \arg\max_{\mathbf{s}} p(\mathbf{s}|\mathbf{y}_t, A_{t-1}) \\ &=\; \arg\max_{\mathbf{s}} p(\mathbf{y}_t|\mathbf{s}, A_{t-1})p(\mathbf{s}) \end{aligned} \tag{3.15}$$

で与えられる。$\hat{\mathbf{s}}_t$ は以下の勾配法を用いて求めた。

$$\triangle\mathbf{s} \propto \frac{1}{\sigma^2}A_{t-1}'(\mathbf{y}_t - A_{t-1}\mathbf{s}) + \frac{\partial}{\partial\mathbf{s}}\log p(\mathbf{s}) \tag{3.16}$$

[1]パラメータ更新が出力値と予測誤差の内積に依存するのは Widrow-Hoff 則 (デルタ則) として知られており、ヘブ学習はその一種である。

この学習則も式 (3.14) と同様にヘブ則の形をとっている。事前分布 $p(\mathbf{s})$ には、

$$\begin{aligned} g(s) &\equiv \frac{\partial}{\partial s}\log p(s) \\ &= -\log\{e^{-a(s-h)} + e^{a(s-h)}\} + b(s-h) + c \end{aligned}$$

を仮定した。ここで、a、 b、 c、h はあらかじめ決められた定数であり、3 節のシミュレーション実験では、$a = 135/7$、$b = 20.7$、 $c = 8.2$、 $h = 0.135$ とした。ここで必要なのは、事前分布そのものではなく対数事前分布の微分であることに注意する。微分 $g(s_i)$ は、i 番目の内部表現 s_i を小さい値にとどめようとし、結果的に内部表現 $\mathbf{s}$ はスパース、すなわち 0 に近い値を取るようになりやすい。こうした事前分布はスパースコーディング事前分布と呼ばれるが、我々のスパースコーディング事前分布は、Olshausen と Field の研究 [11] で用いられたものとは異なる。彼らの事前分布は、我々のより、鋭いピークを持ちかつ裾野が重く、言い換えれば kurtotic なものである。

まとめると、第 3 階層は次のように働く。時刻 t に入力 $\mathbf{y}_t$ が入力されたとき、ネットワークは隠れ変数 $\hat{\mathbf{s}}_t$ の MAP 推定値を式 (3.16) に従って計算する。この推定は、上位レイヤーの認識結果であり、モデル全体の出力として、中枢聴覚システムへ伝送される。同時に、予測値 $A_{t-1}\hat{\mathbf{s}}_t$ が下位レイヤーに戻され、それを用いた誤差 $(\mathbf{y}_t - A_{t-1}\hat{\mathbf{s}}_t)$ が計算される。これらを用いてネットワークのパラメータ A は、式 (3.14) に基づいて更新される。このように、認識と学習が並行して、かつオンライン的に行われる点が特徴である。

3.2. シミュレーション実験と結果

3.2.1 シミュレーション方法

提案モデルの学習実験を実際の音データを用いて行った。訓練データは、ポップミュージックと器楽曲であわせて 22 の楽曲から部分的に抜き出して作成した。音は、サンプリング周波数 44kHz、16 ビット量子化により離散化された。訓練データを連続した 1536 点 (約 35msec) ごとに区分し、それぞれの区間でハニング窓を用いた FFT で周波数分析を行った。さらに要素として周波数ごとの対数パワーをデシベル表示したものをとるベクトルに変換した。計算量の負荷を減らすため、全体で 768 要素あるうち最初の 76 個の低周波成分、すなわち 4362Hz までを用いた。

図 3.4 に、学習後の行列 A を示す。図 3.4 における (i, j) 成分は、 j 番目の内部表現 s_j の活動によってどれだけ i 番目の周波数成分の強度を表現できるかを意味している。つ

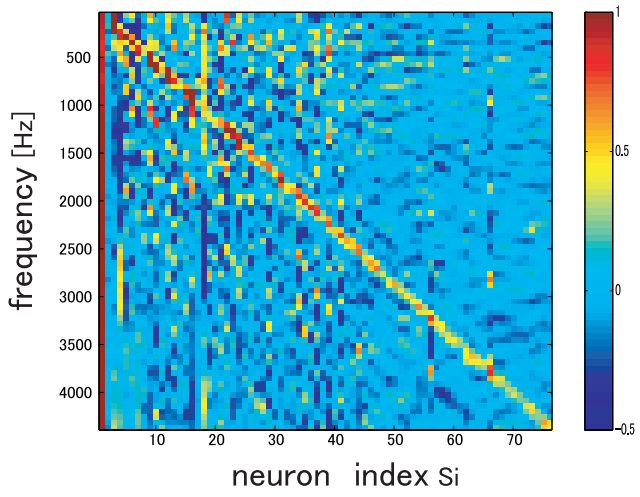


図 3.4 学習後の行列 A。式 (3.6) にあるように、行列 A の各列ベクトルは中枢のニューロンの基底を表す。(i, j) 成分は、j 番目の内部表現 s_j が i 番目の周波数成分とどれだけ関係しているかを示す。右のバーは、重みの値と色づけの対応を示したもので、赤が最も強い重みを表し、青が最も弱い重みであることを示している。

まり、A の j 番目の列ベクトルは、j 番目の内部表現 s_j の基底に対応する。第 3 階層への入力は、内部表現を行列 A で線形変換することで予測される。各々の基底は、どれだけ独立な内部表現が入力周波数情報と関係しているかを示す。図 3.5 に行列 A から 4 つの列ベクトル、すなわちある 4 つの内部表現の基底を抜き出したものを示す。

図のようにそれぞれの基底は、ある基本周波数の整数倍の周波数に強い重みを持っていることがわかる。これらは入力音中の、周波数成分間のハーモニックな構造を取り出していると考えることができる。

3.2.2 Virtual pitch のシミュレーション

2 章で述べたように virtual pitch とは、音に含まれていないはずの周波数ピッチを知覚する錯覚のことである。この virtual pitch が我々のモデルにおいても観測されるかどうかを調べた。図 3.6 は、345Hz の純音をネットワークに入力したときのある内部表現のニューロン $(\mathbf{s}_i)$ の応答と、345Hz 間隔の高調波 1378 Hz、 1723 Hz、 2067 Hz、 2412 Hz の複合音をネットワークに入力したときの応答を比較したもので、どちらも同様にニュー

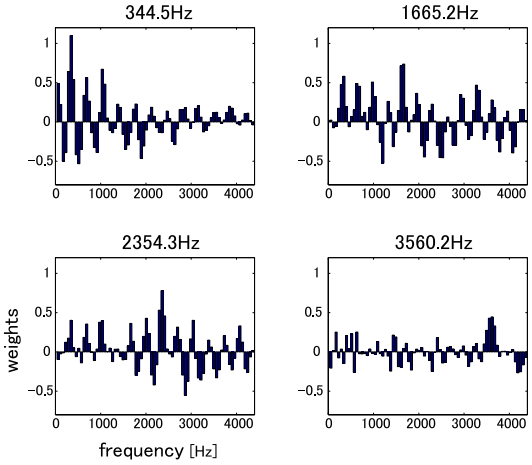


図 3.5　学習後の行列 A から抜き出した 4 つの列ベクトル、すなわち 4 つのニューロンそれぞれがもつ基底ベクトル。タイトルは、そのニューロンが最も強いピークをもつ特徴周波数を表す。いくつかのピークは、特徴周波数だけでなくその約数や倍数のところにも存在する。すなわち基底は、ある基本周波数成分からなるハーモニックな複合音に対して最も強く応答し、ハーモニックな構造を抽出していると考えられる。また、ハーモニックな構造はとりわけ、特徴周波数の低いところで顕著である。

ロンが応答をしていることがわかる。また、高調波複合音のうちどれか単一の周波数だけを与えてもニューロンはあまり応答しなかった。

以上のことから我々は、学習で得られた基底のハーモニックな構造によって virtual pitch が引き起こされると考えた。この解釈は中枢のパターン認識を仮定したモデル [31] [32] [33] と類似するものである。我々のモデルでは、基本周波数が高いときに virtual pitch は観察されなかったが、これは図 3.6 のように基本周波数の高い基底はハーモニックな構造が不明瞭になるためと考えられる。基本周波数が高いと virtual pitch が起りにくいことは、心理学的、生理学的知見とも一致する [43] [52]。

3.2.3 マスキングのシミュレーション

次に、聴神経が伝達すると仮定している、第 3 階層における予測誤差 $(\mathbf{y}_t - A_{t-1}\hat{\mathbf{s}}_t)$ について調べた。マスキングとは、ある刺激を付加することによって他の別の刺激に対する応答が抑圧されたり減少したりする認知的・生理学的現象のことである。生理学的に

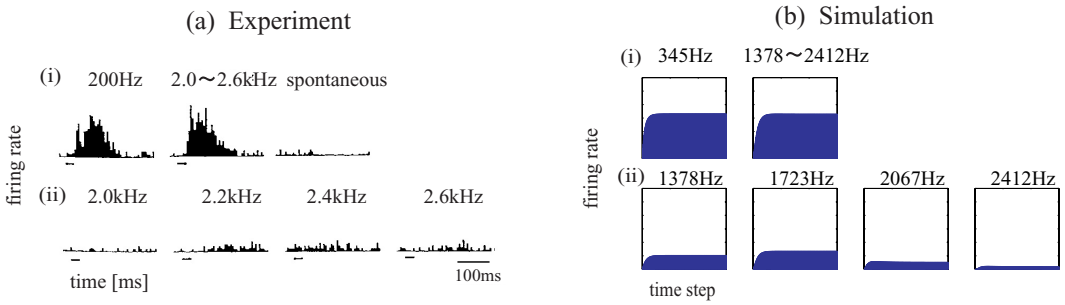


図 3.6 実際の実験で観測された聴覚一次野の神経細胞 (図 2.9 で前出) と、ここでモデル化した第 3 階層のニューロンの純音や複合音に対する応答 (i)、および、複合音のうちの単一の周波数成分のみを聞かせたときの応答 (ii)。(a) 電気生理学的実験 [43]。2000Hz、2200Hz、2400Hz、2600Hz の音を同時に聞かせたとき、サルの聴覚一次野のある神経細胞は、基本周波数の 200Hz の純音を聞かせたときと同様に応答する。しかしながら、その高調波複合音のうちの 1 つの周波数成分だけを聞かせても応答は小さい。(b) 我々のモデルでの数値シミュレーション。ある内部表現 s_i の応答を示している。1378 Hz、 1723 Hz、 2067 Hz、 2412 Hz の音を同時にネットワークに入力したとき、あるモデルニューロンは、基本周波数の 345Hz の純音を聞かせたときと同様に応答する。しかしながら、その高調波複合音のうちの 1 つの周波数成分だけを聞かせても応答は小さい。

は、背景ノイズの付加によって純音に対する聴神経の応答が弱められることが知られている [53]。

図 3.7(a) は、ネコの聴神経におけるマスキング現象を示している。横軸は聞かせる純音の強度を示し、縦軸は聴神経の発火率を示している。複数の線は、異なる背景ノイズの強度における応答関数を示すものである。この図から、背景ノイズの強度が強まったとき、1) 応答関数のベースラインが上昇する、2) 応答関数は右にシフトする、3) 応答関数の飽和値が小さくなる、ということがわかる [53]。これらの生理学的現象は、図 3.7(b) のように学習後の我々のモデルにおいても観察された。

我々のモデルを用いて、マスキング現象は次のように解釈することができる。自然な環境においては純音が単独で出現することはあまりないので、背景ノイズの中に純音が存在するほうがより自然な状況となる。我々のモデルにおいては、この自然さを反映することで背景ノイズが、予測誤差を小さくする効果、すなわち聴神経の発火率を小さくする効果を生んでいると考えることができる。これが応答関数をシフトし、飽和値を下

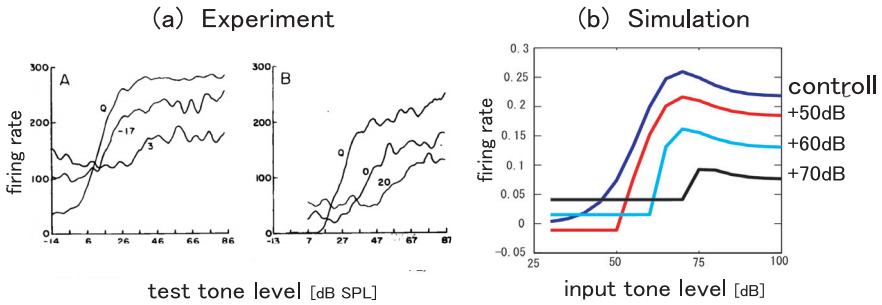


図 3.7 マスキングの効果。実際の実験で観測された聴神経と、ここでモデル化した聴神経とで、背景ノイズの強さを変えたときの純音に対する応答。横軸は、純音の強さを示し、縦軸は、それに対する聴神経の活動を示している。いくつかの曲線は、異なる背景ノイズの強度での応答関数を示している。(a) 2 つの異なる神経細胞に対する電気生理学的実験 [53]。この実験では、背景ノイズの存在下で純音を聞かせている。縦軸は、神経細胞の発火率である。(b) 我々のモデルによる数値シミュレーション。縦軸は、第 3 階層のフィードフォワード信号 $(\mathbf{y}_t - A_{t-1}\hat{\mathbf{s}}_t)$ を示したものである。

げる理由である。ベースラインの上昇は、背景ノイズの付与によりニューロンの活動が全体的に高まるためと考えられる。

3.3. まとめ

これまでに、ピッチ感覚を説明しようするモデルはいくつか提案されている [31] [32] [33] [54]。これらの従来のモデルと我々のモデルの最も大きな違いは、学習可能性である。従来のモデルでは、決定論的な一方向の情報処理システムを仮定し、そこで用いられるパラメータは心理学的現象にあうように設定している。本章では、心理学的現象が双方向のネットワークをもつ情報処理システムにより実現され、そこで用いられるパラメータが情報理論に基づいた学習アルゴリズムにより、経験的に獲得され得ることを示した。実際、発達過程で連続的な雑音環境においてラットの乳児を養育すると、ニューロンのチューニングと全体のサイズという点で聴覚皮質が未成熟なままになってしまう影響がある [55]。

Rao と Ballard のモデル [12] のように我々のモデルでもフィードバック結合を持つ双

方向ネットワークを仮定しており、そのフィードバック結合はオリーブ蝸牛束が担っているとした。麻酔薬アトロピンの注入によってオリーブ蝸牛束を抑制したとき、背景ノイズ下での純音を識別するための最小限の純音強度は、アトロピン注入前と比べて大きくなり、その大きさは背景ノイズの帯域幅が大きいほど顕著になる [56][57]。この知見は、オリーブ蝸牛束が背景ノイズ下で信号を識別するのに重要な役割をはたしていることを示唆している。我々は、オリーブ蝸牛束は新奇な音に対する感度を保持するために予測可能な音を抑制していると考えている。これは、予測可能な情報を除くことで伝達される情報量を大きくする効果がある。さらに、オリーブ蝸牛束は音にさらされたとき、その音に依存した長期可塑性をもつこともわかってきた [58]。このような生物学的知見は、我々のモデルにおけるフィードバック結合の役割と定性的に一致する。

第4章　非線形ノイズ付き独立成分分析による音高知覚モデル

本章では、前章の解剖学的知見を取り入れた 2 段階のエントロピー最大化モデルを非線形ノイズ付き独立成分分析 (ICA) として定式化する。このモデルにおいてもピッチ知覚のシミュレーションが可能であることを示す。

4.1. 数理モデル

4.1.1 確率的エントロピー最大化

ICA の問題は、n 次元観測信号 $\mathbf{x} = [x_1, \cdots, x_n]'$ が n 次元の独立な信号 $\mathbf{s}$ の線形和で表されると仮定したときに、観測信号 $\mathbf{x}$ から元の独立な信号 $\mathbf{s}$ を復元するといった問題である。混合行列を A とすると、この仮定は次の状況を意味している。

$$\mathbf{x} = A\mathbf{s} \tag{4.1}$$

元の独立な信号 $\mathbf{s}$ は、A の逆行列 $W \equiv A^{-1}$ を用いれば、

$$\mathbf{s} = W\mathbf{x} \tag{4.2}$$

で復元可能である。しかし、この復元は行列 A についてはもちろん、$p(\mathbf{s})$ についても、$\mathbf{s}$ の成分が互いに独立であるという知識以外、利用できないとしている。多くの ICA のアルゴリズムは、暫定的に求めた W を用いて元の信号 $\mathbf{s}$ を $\mathbf{y} = W\mathbf{x}$ と推定し、推定された $\mathbf{y}$ がより独立となるようパラメータ W を更新する。独立であるかどうかにの基準には、Kullback-Leibler divergence や結合エントロピー、kurtosis などが用いられる。$p(\mathbf{s})$ が未知であると Kullback-Leibler divergence や結合エントロピーは評価できないが、$p(\mathbf{s})$ に適当な分布を仮定して推定することがよく行われる。また、$p(\mathbf{s})$ に適当な分布を仮定し、最尤推定で直接解くことも多い。本来、$p(\mathbf{s})$ についての知識はないとしているので、$p(\mathbf{s})$

についての仮定によって得られる答えが変わってしまうと問題である。しかし、セミパラメトリック問題を解く推定関数法において、$p(\mathbf{s})$ を適当に仮定したとしても有効な推定関数が構成できることが知られているので、最適解の独立性が保証される [59] [1]。

以上のノイズなし線形 (正方)ICA の定式化は単純であるが、単純であるがゆえにモデルの表現能力が限られている。実際に得られるデータは、

$$\mathbf{x} = A\mathbf{s} + \mathbf{n} \tag{4.3}$$

のようにノイズ $\mathbf{n}$ の存在を仮定したほうが、観測データの生成モデルとしては合理的であることが多い。しかし、このときたとえ W が真の値と一致していても W によって変換された $\mathbf{y}$ は独立とはならない。そのため、ノイズが含まれる場合、観測信号 $\mathbf{x}$ の W による変換で得られた $\mathbf{y}$ を独立化しようとする前述の方法は利用できないことがわかる。また、観測データの信号の数 n が、そこに含まれる独立信号の数 n と一致するという仮定も、想定される状況に応じて変えられるほうが都合がよい。

以上の問題点を克服するために、本章では確率的エントロピー最大化を考える。確率的エントロピー最大化は、決定論的変換によって得られる変数のエントロピーを最大化する決定論的エントロピー最大化に対して、確率的にサンプリングされた変数のエントロピーを最大化するというものである。n 次元観測変数 $\mathbf{x}$ に対して m 次元の隠れ変数 $\mathbf{y}$ をもつ確率モデル $p(\mathbf{x}|\theta)$ を次のように仮定する。

$$p(\mathbf{x}|\theta) = \int p(\mathbf{y})p(\mathbf{x}|\mathbf{y},\theta)d\mathbf{y} \tag{4.4}$$

ただし、隠れ変数に対する事前分布 $p(\mathbf{y})$ は、定義域 m 次元超立方体 $(0,1)^m$ において一様分布するものとする。さらに、観測 $\mathbf{x}$ は $\mathbf{y}$ からのガウス雑音つきの変換によって得られるとする。

$$p(\mathbf{x}|\mathbf{y},\theta) = \frac{1}{Z(\theta)}\exp(-\beta\left\|\mathbf{x} - \phi(\mathbf{y};\theta)\right\|^2) \tag{4.5}$$

ここで、$Z(\theta)$ は積分 $\int p(\mathbf{x}|\mathbf{y},\theta)d\mathbf{x}$ を 1 にする正規化項である。$\beta > 0$ はある定数、$\phi(\mathbf{y};\theta)$ は θ でパラメタライズされる基底関数である。確率的エントロピー最大化は、分布 $p(\mathbf{y}|\mathbf{x},\theta) = \frac{p(\mathbf{y})p(\mathbf{x}|\mathbf{y},\theta)}{p(\mathbf{x}|\theta)}$ に従ってサンプリングされた確率変数 $\mathbf{y}$ のエントロピーを最大化することと定義される。もし、 モデル分布 $p(\mathbf{x}|\theta)$ が真の分布 $q(\mathbf{x})$ をよく近似しているなら $\mathbf{y}$ の分布 $q(\mathbf{y})$ は、 $q(\mathbf{y}) \equiv \int q(\mathbf{x})p(\mathbf{y}|\mathbf{x},\theta)d\mathbf{x} = p(\mathbf{y})\int q(\mathbf{x})\frac{p(\mathbf{x}|\mathbf{y},\theta)}{p(\mathbf{x}|\theta)}d\mathbf{x} \approx p(\mathbf{y})$ の

[1]推定関数を 0 にするパラメータは真のパラメータであることが保証されるが、その推定関数を基にした勾配法で真のパラメータに収束するとは限らない。収束性や収束の早さについての議論は Amari らが行っている [60]。

ように近似される。$p(\mathbf{y})$ は一様分布であるので $q(\mathbf{y})$ も一様分布となり、$q(\mathbf{y})$ のエントロピーは $\mathbf{y}$ の定義域がある有限の値であれば最大化されることがわかる。従って、確率的エントロピー最大化のためには、モデル分布 $p(\mathbf{x}|\theta)$ が真の分布 $q(\mathbf{x})$ をよく近似することが必要であるので、最尤推定を行えばよいことがわかる。このように確率的エントロピー最大化と最尤推定が類似したコスト関数を与えていることがわかる。

次に確率的エントロピー最大化と決定論的エントロピー最大化の関係について述べる。ϕ が逆関数 $\psi \equiv \phi^{-1}$ をもつと仮定する。確率変数 $\mathbf{u} \equiv \phi(\mathbf{y})$ を定義すると、$\mathbf{u}$ と $\mathbf{y}$ との間には以下の関係が成り立つ。

$$d\mathbf{y} = |J(\mathbf{u})| d\mathbf{u}, \tag{4.6}$$

$$\mathbf{y} = \psi(\mathbf{u}), \tag{4.7}$$

ここで、 $|J(\mathbf{u})|$ はヤコビアン $J(\mathbf{u}) \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{u}}$ の絶対値を表す。このとき式 (4.5) は、

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{Z(\theta)} \exp\left(-\beta(\mathbf{x}-\mathbf{u})^T \Sigma^{-1}(\mathbf{x}-\mathbf{u})\right) \equiv p(\mathbf{u}|\mathbf{x}). \tag{4.8}$$

と変数変換できる。関数 $p(\mathbf{u}|\mathbf{x})$ は、$\int p(\mathbf{u}|\mathbf{x}) d\mathbf{u} = 1$ でかつ $p(\mathbf{u}|\mathbf{x}) > 0$ を満たすので確率密度分布であると考えることができる。分布 $p(\mathbf{u}|\mathbf{x})$ は、β が無限大になるときデルタ関数に近づく。

$$\lim_{\beta \to \infty} p(\mathbf{u}|\mathbf{x}) = \delta(\mathbf{u}-\mathbf{x}) \tag{4.9}$$

すなわち、ϕ が逆関数 $\psi \equiv \phi^{-1}$ をもち、β が無限大になるとき、分布 $p(\mathbf{y}|\mathbf{x}, \theta)$ に従った確率的な $\mathbf{y}$ のサンプリングは、$\mathbf{y} = \psi(\mathbf{x})$ の決定論的変換を行うことと一致する。このように、確率的エントロピー最大化の概念は、決定論的なエントロピー最大化の自然な拡張とみなすことができる。

式 (4.6)、 (4.7)、 (4.8) を 式 (4.4) に代入すると、

$$p(\mathbf{x}|\theta) = |J(\mathbf{x})| \tag{4.10}$$

となる。ここで、$J(\mathbf{x}) \equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ であり、$\mathbf{y} = \psi(\mathbf{x})$ である。変換された変数 $\mathbf{y} = \psi(\mathbf{x}; \theta)$ のエントロピーは $H(\mathbf{y}) = H(\mathbf{x}) + \int q(\mathbf{x}) \log |J(\mathbf{x}; \theta)| \, d\mathbf{x}$ で与えられるので、式 (4.10) の最尤推定、$\underset{\theta}{\text{maximize}} \int q(\mathbf{x}) \log |J(\mathbf{x}; \theta)| d\mathbf{x}$ は $\mathbf{y}$ のエントロピー最大化と一致することになる。Cardoso は、決定論的エントロピー最大化と最尤推定が特に、$\psi(\mathbf{x})$ が $n \times n$ の正方行列 W と要素ごとの単調増加関数 $f(\mathbf{x})$ を用いた $\psi(\mathbf{x}) = f(W\mathbf{x})$ のような決定論的変換のときに一致することを示している [61] が、我々の結果はより一般的な仮定である、ϕ が

逆関数 $\psi \equiv \phi^{-1}$ をもつという仮定の下での決定論的エントロピー最大化と最尤推定の等価性を示している。ノイズなし線形 ICA の問題設定において、決定論的変換 $\mathbf{y} = f(W\mathbf{x})$ によるエントロピー最大化は、要素ごとの単調増加関数 f が、それぞれ確率変数 $\mathbf{s} = W\mathbf{x}$ の周辺分布 $p(s_i)$ の累積密度関数となっているとき、$p(\mathbf{s})$ と $\prod_{i=1}^{n} p(s_i)$ の Kullback-Leibler divergence の最小化と一致する。単調増加関数 f が確率変数 $\mathbf{s} = W\mathbf{x}$ の周辺分布の累積密度関数となっていない場合、エントロピー最大化と Kullback-Leibler divergence の最小化は異なるコスト関数となるが、Kullback-Leibler divergence を最小化する変換 W は、エントロピー最大化の局所解となっている [62]。

スパースコーディングモデル [11][63] は、ICA と同様に一次視覚野でみられる線分抽出ニューロンの発現をシミュレートすることができるが、スパースコーディングモデルで用いられる確率モデルは、式 (4.4)、(4.5) の確率モデルと等価であることを示すことができる。スパースコーディングモデルは以下で表される。

$$p(\mathbf{x}|A) = \int p(\mathbf{s})p(\mathbf{x}|\mathbf{s}, A)d\mathbf{s}, \tag{4.11}$$

$$p(\mathbf{x}|\mathbf{s}, A) = \frac{1}{Z}\exp(-\beta \left\|\mathbf{x} - A\mathbf{s}\right\|^2) \tag{4.12}$$

ただし、$\mathbf{s}$ は m 次元の内部表現である。$\mathbf{s}$ の事前分布 $p(\mathbf{s})$ は、0 に鋭いピークをもち、要素間が独立、$p(\mathbf{s}) = \prod_{i=1}^{m} p_i(s_i)$ なスパース事前分布である。A は $n \times m$ の行列である。$g_i(s_i)$ を $p_i(s_i)$ の累積密度関数とし、$g(\mathbf{s}) = [g_1(s_1), \cdots, g_n(s_n)]'$ をそのベクトル表記とする。確率変数 $\mathbf{y}$ を $\mathbf{y} = g(\mathbf{s})$ とおき、関数 f を g の逆関数とする。$\mathbf{s}$ を $\mathbf{s} = f(\mathbf{y})$ とすると、$\mathbf{y}$ の確率分布は $(0, 1)^m$ の一様分布となり、式 (4.12) は式 (4.5) において $\phi(\mathbf{y}; \theta) = Af(\mathbf{y})$ とすると一致する。

4.1.2 非線形ノイズ付き独立成分分析

提案する非線形ノイズ付き ICA のモデルを示す。$\mathbf{x}$ を n 次元観測変数とし、 $\mathbf{s}$ をスパースな分布 $p(\mathbf{s})$ をもつ m 次元隠れ変数とする。ここでは、under-complete な状況、すなわち、n は m と同じかそれより大きい状況を考える。確率モデルは以下のように定式化される。

$$p(\mathbf{x}|A, c) = \int p(\mathbf{s})p(\mathbf{x}|\mathbf{s}, A, c)d\mathbf{s}, \tag{4.13}$$

$$p(\mathbf{x}|\mathbf{s}, A, c) = \frac{1}{Z(A, c)}\exp\left(-\beta \left(g(\mathbf{x}; c) - A\mathbf{s}\right)' \Sigma(\mathbf{x}; c) \left(g(\mathbf{x}; c) - A\mathbf{s}\right)\right), \tag{4.14}$$

$$p(\mathbf{s}) = \prod_{i=1}^{m} \frac{1}{2} \left(1 - \tanh^2(s_i)\right) \tag{4.15}$$

ここで、A は $n \times m$ 行列、 $g(\mathbf{x}; c)$ は $g(x_i; c_i) = \log(x_i) + c_i$ である。$\Sigma(\mathbf{x})$ は (i, i) 成分が $[\Sigma(\mathbf{x})]_{ii} = x_i^2$ である $n \times n$ の対角行列で、$Z(A, c)$ は積分 $\int p(\mathbf{x}|\mathbf{s}, c, A)d\mathbf{x}$ を 1 にする正規化定数である。パラメータは、A と c である。

この確率モデル $p(\mathbf{x}|\mathbf{s}, c, A)$ は後段に非線形関数を含めた ICA の定式化 $\mathbf{x} = f(A\mathbf{s}) + \mathbf{n}$ に基づいている。$f(\cdot)$ は $g(\cdot)$ の逆関数 とする。$\mathbf{n}$ は平均 0、共分散が $(2\beta)^{-1}I$ の対角行列である。I は単位行列を表す。もしも、$(\mathbf{x} - f(A\mathbf{s}))$ が 0 に十分近ければ、$\mathbf{x} - f(A\mathbf{s}) \approx f'(g(\mathbf{x}))\,(g(\mathbf{x}) - A\mathbf{s})$ とかける。ここで、 $f'(\mathbf{x}; c)$ は (i, i) 成分が $[f'(\mathbf{x}; c)]_{ii} = \frac{\partial f(x_i; c_i)}{\partial x_i}$ である対角行列である。$[\Sigma(\mathbf{x}; c)]_{ii} = \{f'(g(x_i; c); c)\}^2$ とすると、$\mathbf{x}$ の $\mathbf{s}$ が与えられた下での条件付き分布は式 (4.14) で表される。

一般化 EM アルゴリズム

隠れ変数 $\mathbf{s}$ を含むモデルの最尤推定となるので、パラメータ A と c についての対数尤度 (4.13) の最大化のために EM アルゴリズム [49] を用いた。

- E ステップ

  現在のパラメータ推定値 $\bar{A}$ と $\bar{c}$ を用いて期待対数尤度を計算する。

  $$Q(A, c|\bar{A}, \bar{c}) = \frac{1}{T} \sum_{t=1}^{T} \int p(\mathbf{s}_t|\mathbf{x}_t, \bar{A}, \bar{c}) \log p(\mathbf{x}_t, \mathbf{s}_t|A, c) d\mathbf{s}_t \tag{4.16}$$

  ここで、$\{\mathbf{x}_t\}$ は T 個のサンプル集合を表し、$\{\mathbf{s}_t\}$ は対応する隠れ変数を表す。

- M ステップ

  パラメータ A と c について $Q(A, c|\bar{A}, \bar{c})$ を最大化する。.

E-step

$\log p(\mathbf{x}, \mathbf{s}|A, c) = -\beta\,(g(\mathbf{x}; c) - A\mathbf{s})'\,\Sigma(\mathbf{x}; c)\,(g(\mathbf{x}; c) - A\mathbf{s}) - \log Z(A, c) - \log p(\mathbf{s})$ のパラメータに依存する項は、隠れ変数 $\mathbf{s}$ について 1 次と 2 次の統計量しか含んでいないので、式 (4.16) の積分はラプラス近似によって計算できる。

ラプラス近似によれば、式 (4.16) に現われる $p(\mathbf{s}|\mathbf{x}, \bar{A}, \bar{c})$ は、最大事後確率をとる MAP 値 $\hat{\mathbf{s}} \equiv \arg\max p(\mathbf{s}|\mathbf{x}, \bar{A}, \bar{c}) = \arg\max_{\mathbf{s}} \{\log p(\mathbf{x}|\mathbf{s}, \bar{A}, \bar{c}) + \log p(\mathbf{s})\}$ においてピークをもつガウス分布に近似される。

$$p(\mathbf{s}|\mathbf{x}, \bar{A}, \bar{c}) \propto \exp\left(-\frac{1}{2}(\mathbf{s}-\hat{\mathbf{s}})' H(\hat{\mathbf{s}})(\mathbf{s}-\hat{\mathbf{s}})\right) \tag{4.17}$$

ただし、 $H(\hat{\mathbf{s}}) \equiv -\nabla\nabla \log p(\hat{\mathbf{s}}|\mathbf{x}, \bar{A}, \bar{c})$ である。

MAP 値 $\hat{\mathbf{s}}$ は、勾配法によって求められる。

$$\Delta\mathbf{s} \propto 2\beta \bar{A}' \bar{\Sigma} \left(g(\mathbf{x}; \bar{c}) - \bar{A}\mathbf{s}\right) + \varphi(\mathbf{s}) \tag{4.18}$$

ここで、 $\varphi(\mathbf{s}) \equiv \frac{\partial \log p(\mathbf{s})}{\partial \mathbf{s}} = -2\tanh(\mathbf{s})$ である。

これより、1 次と 2 次のモーメントは

$$\langle \mathbf{s} \rangle \equiv \int p(\mathbf{s}|\mathbf{x}, \bar{A}, \bar{c}) \mathbf{s} d\mathbf{s} = \hat{\mathbf{s}}, \tag{4.19}$$

$$\langle \mathbf{s}'\mathbf{s} \rangle \equiv \int p(\mathbf{s}|\mathbf{x}, \bar{A}, \bar{c}) \mathbf{s}\mathbf{s}' d\mathbf{s} = \hat{\mathbf{s}}\hat{\mathbf{s}}' + H(\hat{\mathbf{s}})^{-1}, \tag{4.20}$$

で求められる。また、 $H(\hat{\mathbf{s}})^{-1}$ は β が十分大きいとする低ノイズ近似を行ったとき

$$\begin{aligned} H(\hat{\mathbf{s}})^{-1} &= \left(2\beta A'\Sigma A - \nabla\nabla \log p(\hat{\mathbf{s}})\right)^{-1}, \\ &\approx \frac{1}{2\beta}\left(A'\Sigma A\right)^{-1} + \frac{1}{4\beta^2}\left(A'\Sigma A\right)^{-1} \nabla\nabla \log p(\hat{\mathbf{s}}) \left(A'\Sigma A\right)^{-1} \end{aligned} \tag{4.21}$$

で計算できる。

M-step

式 (4.19) と式 (4.20) を用い、正規化項 $(-\log Z(A, c))$ を低ノイズ近似によって無視すると、期待対数尤度 $Q(A, c|\bar{A}, \bar{c})$ の微分は以下のように計算される。

$$\frac{\partial Q}{\partial A} \approx \frac{2\beta}{T} \sum_{t=1}^{T} \left(\Sigma_t \mathbf{e}_t \hat{\mathbf{s}}_t' - \Sigma_t A H(\hat{\mathbf{s}}_t)^{-1}\right), \tag{4.22}$$

$$\frac{\partial Q}{\partial c_i} \approx -\frac{2\beta}{T} \sum_{t=1}^{T} e_{t,i} f'(g(x_{t,i})) \left(f'(g(x_{t,i})) + e_{t,i} \frac{\partial f'(g(x_{t,i}))}{\partial c_i}\right) \tag{4.23}$$

ただし、 $\mathbf{e}_t \equiv (g(\mathbf{x}_t; c) - A\hat{\mathbf{s}}_t)$ である。また、$e_{t,i}$ は $\mathbf{e}_t$ の i 番目の要素であり、$x_{t,i}$ は $\mathbf{x}_t$ の i 番目の要素である。式 (4.23) については、$\frac{\partial Q}{\partial c} = 0$ の解析解が求められ、解析解は次

のようになる。

$$c_i = -\frac{\sum_{t=1}^{T} x_{i,t}^2 \left(\log x_{i,t} - \sum_{j=1}^{m} a_{ij} \hat{s}_{j,t} \right)}{\sum_{t=1}^{T} x_{i,t}^2} \tag{4.24}$$

このモデルの場合、パラメータ A についての解析解は得られないので M ステップは勾配法に基づいて期待対数尤度を向上させることになり、EM アルゴリズムの手続きは一般化 EM アルゴリズムとなる。もしも、Σ_t が定数、すなわち、モデルが線形なノイズつき ICA であれば式 (4.22) において $\frac{\partial Q}{\partial A} = 0$ の解析解が得られる。

$$A = \frac{1}{T} \sum_{t=1}^{T} g(\mathbf{x}_t; c) \hat{\mathbf{s}}_t' \left\{ \frac{1}{T} \sum_{t=1}^{T} \left(\hat{\mathbf{s}}_t \hat{\mathbf{s}}_t' + H(\hat{\mathbf{s}}_t)^{-1} \right) \right\}^{-1} \tag{4.25}$$

これは、Bermond と Cardoso のノイズ付き ICA[64] に一致する。非線形 ICA モデルの場合は M ステップ解析解は得られず、勾配法によるパラメータの更新を行う。ここでは、自然勾配法を用いてパラメータ A の学習則を導出した。パラメータのユークリッド距離に基づく勾配 $\frac{\partial Q}{\partial A}$ の代わりに、自然勾配 $AA' \frac{\partial Q}{\partial A}$ を ICA では用いる。この勾配を用いたほうが実用的には学習が早くなることが示されている [63][65]。

$$AA' \frac{\partial Q}{\partial A} \approx \frac{2\beta}{T} \sum_{t=1}^{T} \left(AA' \Sigma \left(g(\mathbf{x}_t; c) - A \hat{\mathbf{s}}_t \right) \hat{\mathbf{s}}_t' - AA' \Sigma A H(\hat{\mathbf{s}}_t)^{-1} \right)$$

MAP 値 $\hat{\mathbf{s}}$ は $\Delta \mathbf{s} = 0$、 すなわち、 $2\beta A' \Sigma \left(g(\mathbf{x}; c) - A\mathbf{s} \right) + \varphi(\mathbf{s}) = 0$ を満たす必要があるので以下の変形が可能になる。

$$AA^T \frac{\partial Q}{\partial A} \approx -\frac{1}{T} \sum_{t=1}^{T} A \left(\phi(\hat{\mathbf{s}}_t) \hat{\mathbf{s}}_t^T + 2\beta A^T \Sigma_t A H(\hat{\mathbf{s}}_t)^{-1} \right). \tag{4.26}$$

ノイズが十分小さい、すなわち β が十分大きいと仮定できるとき、$H(\hat{\mathbf{s}})^{-1} \approx \frac{1}{2\beta} \left(A' \Sigma A \right)^{-1} + \frac{1}{4\beta^2} \left(A' \Sigma A \right)^{-1} \nabla \nabla \log p(\hat{\mathbf{s}}) \left(A' \Sigma A \right)^{-1}$ と近似が成り立つので

$$AA' \frac{\partial Q}{\partial A} \approx -\frac{1}{T} \sum_{t=1}^{T} A \left(I + \varphi(\hat{\mathbf{s}}_t) \hat{\mathbf{s}}_t' + \frac{1}{2\beta} \nabla \nabla \log p(\hat{\mathbf{s}}_t) \left(A' \Sigma A \right)^{-1} \right) \tag{4.27}$$

となる。特に、A が逆行列 $W \equiv A^{-1}$ をもつとき、$\frac{\partial Q}{\partial W} = -A' \frac{\partial Q}{\partial A} A'$ であることに注意すると、式 (4.27) の自然勾配法は

$$\frac{\partial Q}{\partial W} W'W \quad = \quad -A' \frac{\partial Q}{\partial A} A'W'W$$

$$= \frac{1}{T}\sum_{t=1}^{T}\left(I + \phi(\hat{\mathbf{s}}_t)\hat{\mathbf{s}}_t' + \frac{1}{2\beta}\nabla\nabla \log p(\hat{\mathbf{s}}_t) W\Sigma^{-1}W'\right)W \qquad (4.28)$$

とかける。これは、Douglas、 Cichocki と Amari[66] の低ノイズ近似を用いたノイズ付き線形 ICA と非常に近いアルゴリズムとなっている。私が $\hat{\mathbf{s}} = \arg\max_{\mathbf{s}} p(\mathbf{s}|\mathbf{x}) = \arg\max_{\mathbf{s}}[p(\mathbf{x}|\mathbf{s})p(\mathbf{s})]$ としているのに対して、彼らのアルゴリズムでは $\hat{\mathbf{s}} = \arg\max_{\mathbf{s}} p(\mathbf{x}|\mathbf{s}) = W\mathbf{x}$ であり、また式 (4.28) の右辺第 2 項の符号が彼らのアルゴリズムと逆になっている。これら 2 つのノイズ付き線形 ICA アルゴリズムの重要な点は、ノイズがなくなる極限、すなわち β が無限大となる極限において右辺第 2 項が 0 に近づきノイズなし ICA アルゴリズムと一致することである。Lewicki と Sejnowski の提案したノイズ付き ICA アルゴリズム [63] を除いては、これまでの最尤推定に基づいて導出されたノイズ付き ICA アルゴリズム [51][67] [64][68] は積分の近似計算が正確でなかったことにより、ノイズのなくなる極限ではノイズなし ICA アルゴリズムとは異なったものとなっていた。ただし、Lewicki と Sejnowski は、尤度を計算する積分 $\int p(\mathbf{s})p(\mathbf{x}|\mathbf{s},c)d\mathbf{s}$ に関与しないと考えて無視している項があり、その項を消去したときにノイズなし ICA アルゴリズムと一致する。またその項は、学習を不安定にさせるとも報告している。それに対して本章で提案したアルゴリズムでは、ノイズのなくなる極限において右辺第 2 項は自動的に 0 になることでノイズなし ICA アルゴリズムと一致する。本章の近似計算も Lewicki と Sejnowski の近似計算も MAP 値でガウス近似を行うものであったが、近似計算する対象が異なっている。私が、MAP 値 $\hat{\mathbf{s}}$ と無関係にパラメータ A や c の最適化ができる EM アルゴリズムにおける期待対数尤度 $Q(A,c|\bar{A},\bar{c})$ に含まれる積分計算を近似したのに対し、Lewicki と Sejnowski は MAP 値 $\hat{\mathbf{s}}$ がパラメータ A や c の最適化と密接に関係する尤度そのもの $\int p(\mathbf{s})p(\mathbf{x}|\mathbf{s},c)d\mathbf{s}$ の計算に対して近似計算を行っている。パラメータ A を変化させたときに MAP 値 $\hat{\mathbf{s}}$ がどう影響されるかを見積もるのは難しく、その見積もりのためにさらに近似計算を行っているため、学習を不安定にさせるような項が生じたものと考えられる。

また、もし式 (4.22) の項 $\Sigma A H(\hat{\mathbf{s}}_t)^{-1}$ を無視すると、MAP 値での事後確率をデルタ関数近似することになる。このとき、スパースコーディングモデル [11] と一致する。

$$\Delta A \propto \frac{2\beta}{T}\sum_{t=1}^{T}\Sigma_t\left(g(\mathbf{x}_t;\theta) - A\hat{\mathbf{s}}_t\right)\hat{\mathbf{s}}_t^T \qquad (4.29)$$

線形なモデル、すなわち $g(\mathbf{x};c)$ が線形である場合、式 (4.29) における Σ は単位行列となる。スパースコーディングモデルは、始めの定式化ではノイズなし ICA を含むノイズ付き ICA として定式化されているが、最終的に導出される学習則、式 (4.29) はノイズな

し ICA の学習則を含む形とはなっていない。これは、ノイズを許容することで生じた積分の計算を、事後分布が MAP 値においてデルタ関数近似されるとした粗い近似を行ったからであると考えられる。事後分布のデルタ関数近似は、ノイズなし ICA の学習則にバイアスを加えてしまうものであることがわかった。以上のように、ここで提案した非線形ノイズ付き ICA アルゴリズムは、いくつかの線形ノイズ付き ICA アルゴリズムや線形ノイズなし ICA アルゴリズムを含む形になっている。また、この ICA アルゴリズムは決定論的エントロピー最大化に基づく非線形ノイズなし ICA [69] の一般化となっている。なぜなら 4.4.1 節で述べたように決定論的エントロピー最大化は最尤推定と等価なコスト関数であるからである。

4.1.3 ステップサイズの設定

一般化 EM アルゴリズムにより、パラメータ A は勾配法に基づいて更新 ($A := A+\epsilon\Delta A$) される。ここでは、そのときの適切なステップサイズ ϵ の設定方法について議論する。一般化 EM アルゴリズムでは、ステップサイズは期待対数尤度 $Q(A, c|\bar{A}, \bar{c})$、もしくは更新前のパラメータを用いた期待対数尤度との差分 $[Q(A, c|\bar{A}, \bar{c}) - Q(\bar{A}, \bar{c}|\bar{A}, \bar{c})]$ を最大化するように決められるべきである。パラメータ c は、式 (4.24) に基づいて更新されるとき $Q(\bar{A}, c|\bar{A}, \bar{c}) \geq Q(\bar{A}, \bar{c}|\bar{A}, \bar{c})$ を満たすので、パラメータ A は $[Q(A, c|\bar{A}, \bar{c}) - Q(\bar{A}, c|\bar{A}, \bar{c})]$ が最大となるように更新すればよい。そこで、ステップサイズ ϵ についてのコスト関数を次のように定める。

$$F(\epsilon) \equiv Q(A, \theta|\bar{A}, \bar{\theta}) - Q(\bar{A}, \theta|\bar{A}, \bar{\theta}) \tag{4.30}$$

ただし、$A = \bar{A} + \epsilon\Delta A$ である。このコスト関数 $F(\epsilon)$ は、ノイズが十分小さいとみなせるか、非線形関数 $g(\mathbf{x}_t; \theta)$ が線形な場合、単にステップサイズ ϵ についての 2 次形式となる。

$$F(\epsilon) \approx -a\left(\epsilon - \frac{b}{2a}\right)^2 + \frac{b^2}{4a} \tag{4.31}$$

ただし、$a = \frac{\beta}{T}\sum_{t=1}^{T}\left(\mathrm{Tr}\left(\langle \mathbf{s}_t\mathbf{s}_t'\rangle \Delta A'\Sigma_t\Delta A\right)\right)$ であり、$b = \frac{2\beta}{T}\sum_{t=1}^{T}\left(g(\mathbf{x}_t;\theta)'\Sigma_t\Delta A\langle \mathbf{s}_t\rangle - \mathrm{Tr}\left(\langle \mathbf{s}_t\mathbf{s}_t'\rangle \Delta A'\Sigma_t\bar{A}\right)\right)$ である。$\mathrm{Tr}(\cdot)$ は行列のトレースを表す。式 (4.31) の導出は、付録 A.1 に示す。係数 a と b は正であるから (付録 A.2 参照)、最適なステップサイズは

$$\epsilon = \frac{b}{2a}, \tag{4.32}$$

で与えられる。これは与えられた勾配方向において最適なステップサイズであるので、ラインサーチでパラメータ A を更新することに相当する。

ノイズのない極限においては係数 a が発散する一方、係数 b はある有限の値に収束する。これはパラメータ更新による期待対数尤度の増分がノイズが小さくなればなるほど、小さくなり、ノイズが 0 になる極限では増分が 0 になることを示している。これは、ノイズのない極限においては EM アルゴリズムによるパラメータの更新では尤度の増加が見込めなくなることを意味する。しかしながら、ノイズのない極限においても期待対数尤度を増加させるための勾配は意味をもつ。この勾配は、期待対数尤度を増大させることはなくなるが尤度そのものをパラメータで微分した勾配に一致するため、尤度を増大させることができるからである。そのためノイズがかなり小さい状況においては、ここで求めた最適なステップサイズより大きくしても尤度を増大させることができる。

4.2. シミュレーション

4.2.1 シミュレーション方法

シミュレーションに用いたデータは、音声と音楽からなる。音声は、2 名の日本人と 4 名のスペイン人、4 名のアメリカ人の音声をあわせて 232 秒分取得した。音楽は、ポップミュージックとリラクゼーション音楽 (どちらも音声と楽音を含む) のあわせて 25 曲から抜き出したものと、ピアノ曲、ヴァイオリン協奏曲、オーケストラのあわせて 24 曲からそれぞれ 116 秒間のデータを得た。どちらの音源も 44.1kHz のサンプリング周波数で 16 ビットに離散化されたデータである。これを 11025Hz にダウンサンプリングし、512 点ずつ離散フーリエ変換によって周波数領域に変換している。離散フーリエ変換して得られた係数から周波数成分ごとの振幅スペクトルを計算し、これを観測変数 $\mathbf{x}$ とした。$\mathbf{x}$ は 256 次元のスペクトルパターンとなり、全部で 1 万点得られる。MAP 値 $\hat{\mathbf{s}}$ が脳内の内部表現を表すと考え、内部表現 $\mathbf{s}$ の次元 m は、$m = 30$ とした。行列 A は 256×30 の行列となる。

4.2.2 ピッチの存在領域の評価

図 4.1 に学習後の行列 A を示す。図 4.1 は縦軸が n 次元の周波数成分、横軸が m 次元のピッチパターンの尺度に対応した行列 A の各要素の値を色づけして表している。図 4.2

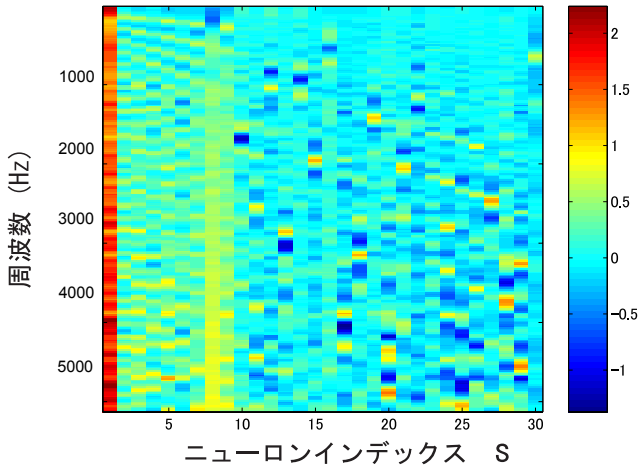


図 4.1　学習した行列 A

にその行列 A の列ベクトルを 4 つ抜き出したものを示す。列ベクトルは、ある内部表現 s_j の基底ベクトルとなるものである。基底ベクトルは、周期的にピークをもつことがわかる。周期的なピークには丸で印がつけられている。そして図上部のタイトルは、その丸で印がつけられた周波数の基本周波数を示している。また、図 4.1 と図 4.2 からわかるように低周波数成分にはそれほど大きな重みをもっていない。図 4.2 の左下の基底ベクトルは周期構造がはっきりしていなかった。この基底ベクトルは、次節で示すようにどの周波数入力に対しても常に大きな値をとる内部表現 s_j に対応しており、音の大きさ、ラウドネスを表現するパラメータであると考えた。

この学習後の行列 A を用いて、virtual pitch のシミュレーションを行った。ピッチは、最も大きな値をとる内部表現、ニューロン s に対応すると仮定した。すなわち、入力 A に対してニューロン集団 **s** の中で最も大きな値、発火率をとるニューロン s_j が入力 B に対しても最も大きな発火率をとっていた場合、入力 A と入力 B は同じピッチを与えていると仮定する。3 章のモデルでは、基本周波数のみからなる純音を与えたときと、基本周波数を含まない高調波複合音を与えたときとで、同じように強く活動するニューロンが見出されていることを示し、これが virtual pitch 知覚の原因となっていると考えたが、ここではさらに一歩踏み込んでどのような入力音に対して virtual pitch が観察されるかを実験と比較した。

2.1.2 節で述べたように Ristma は、基本周波数 g とその 2 次と 3 次の倍音からなる低調波複合音 (簡単のために基本周波数音とする) と、周波数 $(f-g)$、f、$(f+g)$ からな

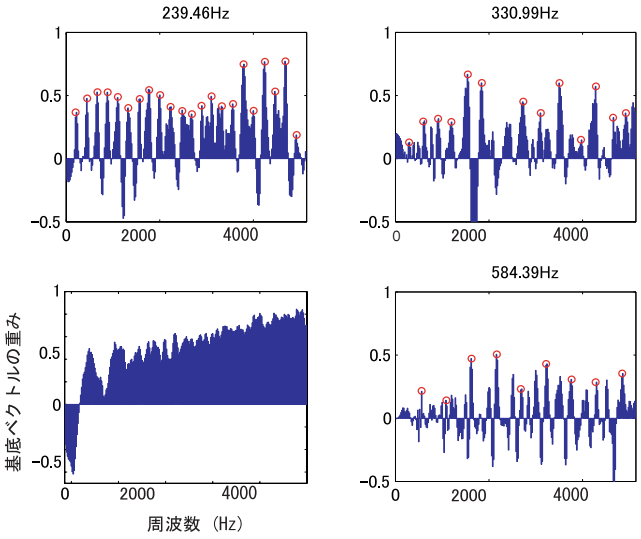


図 4.2 学習した行列 A の基底ベクトル

る高調波複合音とがどのような領域 f, g で同じピッチを与えるかを調べた [22]。図 2.5 と同じ図を図 4.3 に再掲する。

図 4.4 に基本周波数音と高調波複合音を入力したときの内部表現ニューロン $\mathbf{s}$ のとる値、発火率を示す。どの入力に対しても 8 番目のニューロン s_8 が強く応答している。このニューロンの基底ベクトルが、前述の図 4.2 の左下の基底ベクトルに対応している。しかし、その 8 番目のニューロン s_8 を除けば、入力に応じて異なるニューロンが強く応答していることがわかる。また、そのニューロンは基本周波数音と高調波複合音とでほぼ同じである。そのため、ニューロン s_8 は音の大きさ、ラウドネスを表すニューロンと考えてニューロン s_8 を無視した他のニューロンが同じように強く活動したときにのみ同じピッチを与えると判断した。この結果を図 4.5 に示す。図中の丸印は、基本周波数音と高調波複合音のピッチが一致したことを示し、三角印は、基本周波数音に対して最も強く応答したニューロンが高調波複合音に対しては 2 番目に強く応答したことを示している。点は、基本周波数音と高調波複合音のピッチが一致しなかった場所を示す。図のように、図 4.3 の実験と同様に基本周波数が低周波数域と高周波数域にあるときは virtual pitch が観測されないことがわかった。

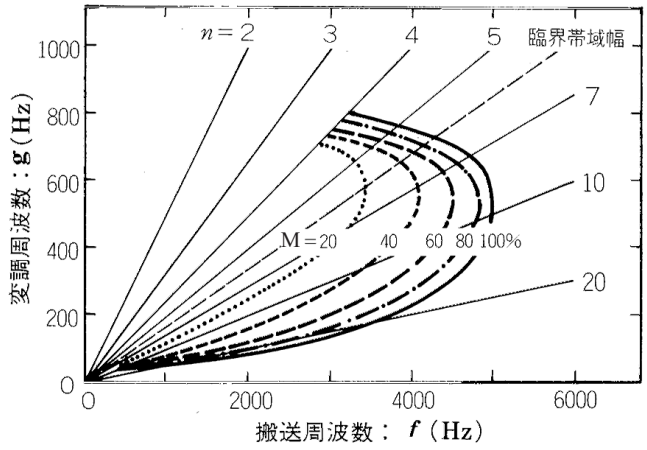


図 4.3 ピッチの存在する領域。Ristma (1962) を基に引用。横軸は高調波複合音のうちの中心周波数 f を表し、縦軸は、基本周波数 g を表す。M は変調度である。(変調度の定義については 2.1.2 節を参照)

4.3. まとめ

本章では、非線形ノイズ付き ICA を提案した。これは、従来のノイズ付き ICA を非線形に一般化するものとなった。提案した非線形ノイズ付き ICA を用いて、音声や音楽のデータを学習した。学習した結果を用いて virtual pitch の存在領域について実験結果と比較したところ、定性的には良い一致を見出せた。また、得られた基底ベクトルから基本周波数の存在する低周波数帯域より高い周波数帯域の情報を利用してピッチ情報を抽出することが示唆された。これは、3 次から 5 次にかけての倍音がピッチの知覚に支配的であるとする実験データと一致する [70][71]。また、2 節で Wightman のモデルがケプストラム分析と近いモデルであることを指摘したが、ここで提案した非線形ノイズ付き ICA は、パラメータの設定によってケプストラム分析と完全に一致させることができる。ケプストラム分析では、行列をフーリエ基底に固定しているが、ピッチを特徴付ける独立な基底を非線形ノイズ付き ICA で学習することによってさらに音声認識の性能を向上できる可能性も考えられる。

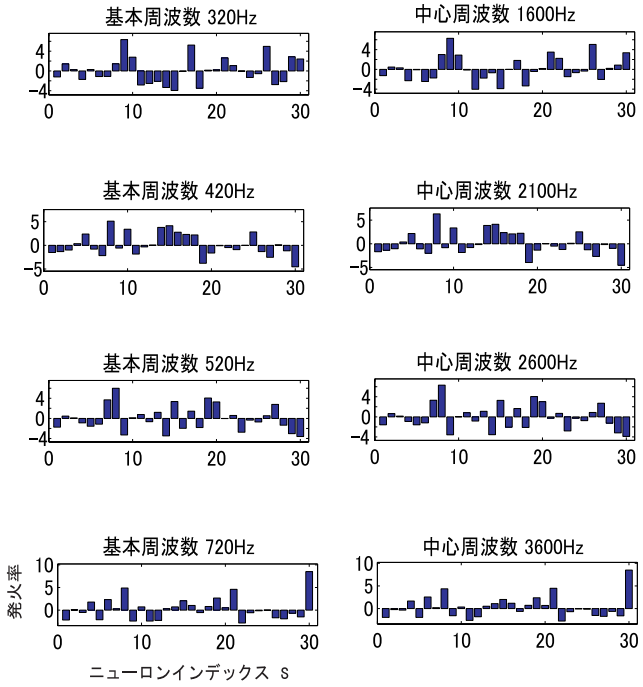


図 4.4 入力が基本周波数音、高調波複合音のそれぞれのときのニューロン s のとる値の大きさ (発火率)

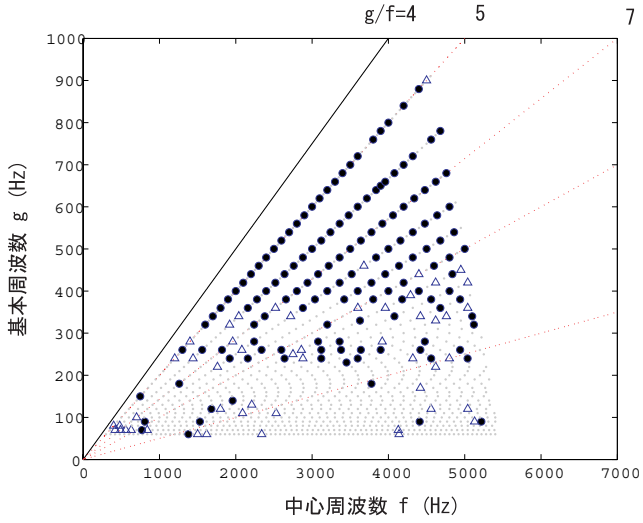


図 4.5 モデルによって模擬されたピッチの存在領域。横軸は高調波複合音のうちの中心周波数 f を表し、縦軸は、基本周波数 g を表す。

第5章　レート歪み理論に基づいた効率的な情報表現

ここでは、エントロピー最大化や独立化に基づく効率的な情報表現をレート歪み理論の立場から説明し、高い表現能力をもった制約付き歪みあり符号である product code の最適化法について述べる。パーセプトロン型の変換関数を product code に用いたとき、変換によって独立化できるような線形変換が学習できることを示し、最後にその理由について考察する。

5.1. 歪みあり符号

5.1.1 脳の情報表現と歪みあり符号の関係

ここまで、エントロピーを最大にする符号化は、情報の性質を予測した冗長性を圧縮した表現をとることで、効率的な符号化になると解釈してきた。しかし、この効率的とは何をもって効率的といえるか曖昧なままの議論であった。そこで、効率性をレート歪み理論の立場で定式化し、直接的に効率的な符号化を行うことを考える。レート歪み理論で定式化するということはリソース (符号長) が限られていることを前提にしている。脳には、100 億〜1 千億ばかりのニューロンがあると言われている。しかし、仮に 1 千億個のニューロン各々が発火する、発火しないの 2 通りの情報を表現できたとしても全体では、2 の 1 千億乗、バイトに換算すると高々12 ギガバイトの情報表現容量でしかない。これは、現在のノートパソコンの記憶容量にも劣る容量である。すなわち、リソースとなる神経細胞を有効に使うためには高効率な圧縮が不可欠である。そこで、レート歪み理論の立場から高性能な圧縮符号について議論を行い、それに基づく脳の情報処理効率の議論への展開、なぜ内部状態が独立化するような基底が得られるかの説明を試みる。

5.1.2 歪みあり符号の概説

歪みあり符号とは、復号した reproduction と符号元の情報源との不一致を許容し、その不一致の度合い (歪み) をある歪み基準によって測ったときに、その歪みが小さくなるようにしつつ、符号化時の符号長をできるだけ小さくするという符号化方法である。ここでは主に符号長を制限したときに、その制約の中で最も歪みを小さくする符号化器を求める問題を考える。符号化したい情報の系列をあるブロックに区切り、ブロック毎に符号化する方法をブロック符号化というが、歪みあり符号の問題ではベクトル量子化がブロック符号化の中では最適[1]な符号化方法を与えるものとして古くから研究がなされてきた [72] [73]。しかしベクトル量子化は符号化時に情報源の次元に対して指数的に計算量が大きくなる問題が指摘されてきた。またベクトル量子化の設計に、一般化された Lloyd アルゴリズム (GLA) などを用いたとき [13] にもやはり符号化時と同様の計算量が必要なので、これはベクトル量子化の設計が困難なことも同時に意味している。そこでベクトル量子化に何らかの制約をかけることで計算量を削減することが考えられている。そもそもベクトル量子化の計算量が情報源の次元に対して指数的に増大するのは、量子化の候補を情報源の次元 n のべき、すなわち R^n の空間から探索する必要があることに起因している。Product code はベクトル量子化をスカラー量子化の集合におきかえるという、制約のある量子化関数を用いた符号化方法である [73] [74]。スカラー量子化の採用によって次元ごとの探索を R^1 の空間に制限することができ、そのため情報源の次元に対してはたかだか線形の計算量にとどめることができる。Product code は計算量において大きなメリットをもつので、その性能を最大限に発揮する符号器は有望である。

簡単な量子化器で高い圧縮性能を引き出す方法として、情報源をその量子化器の特性にあうように変換することが考えられている [75]。Companding 関数はそうした変換を行う関数であり、compressor と expander からなる。Compressor は情報源を変換し、変換後の信号に対して量子化が行われる。そして expander によって量子化された信号が再変換され、これが reproduction となる。簡単な量子化器としてしばしば用いられるのは、ラティス量子化かその特殊な場合である一様量子化である。Product code において、要素ごとのスカラー量子化器は companding 関数を用いることで区間 $(0, 1)$ の一様量子化器として表現できることに注意する。本論文では、要素ごとに区間 $(0, 1)$ での一様量子化器を用いた場合の最適な多次元 companding 関数を求めることを考え、その集合を用いた

[1] ここで「最適」とは、歪みあり符号化の観点のもとでの最適性をさす。またこれ以降もとくに断りがなければ、「最適」は歪みあり符号の観点のもとでの最適性をさす。

product code による符号化方法を提案する。以後本論文では、この符号化方式のことを product code と呼ぶ。

一方、たとえ情報源の分布が明示的に与えられていたとしても最適な companding 関数を求めることは難しい。スカラー量子化においては高量子化の極限において二乗誤差を最小にする companding 関数は導出されている [76] が、ブロック符号に拡張した多次元 companding 関数の導出はブロック内の各要素間が独立でかつ、高量子化の極限で符号化する場合 [77] 以外は知られていない。さらに companding 関数のクラスを限定することでそのクラス内で最適な companding 関数を求めることは可能である。よく知られている結果は、companding 関数を直交線形変換に固定し情報源をガウス分布とした場合で、このとき二乗誤差最小の意味で最適なものは Karhunen-Loeve 変換 (KL 変換) となる [78]。上記のように特別な条件のもとでの最適な companding 関数が解析的に調べられている一方で、そもそも最適な companding 関数が最適なブロック符号化方法、つまりベクトル量子化を表現できるかに関しても議論されている。その結果、大多数の情報源に対して最適なベクトル量子化が多次元 companding 関数をラティス量子化に用いた符号器で表現できないことから、多次元 companding 関数を一様量子化器に用いた符号化方法である product code に限界があることが示唆されている [79] [80]。

情報源の分布が与えられていても明示的な多次元 companding 関数を得ることが困難であったことや、仮に最適な多次元 companding 関数が得られたとしても大抵の場合、最適なブロック符号化方法とはならないこと、離散コサイン変換を用いた変換符号化など計算が高速でかつ高圧縮を可能にする符号化方法が提案されていたこと [81] などの理由により、これまで多次元 companding 関数を最適化するという視点はあまり注目されてこなかった。しかしながら、現在、高圧縮性能を持つ符号化方法として提案されている Tree-Structured Vector Quantization(TSVQ) は、product code とみなすことができる [74]。TSVQ で用いる木構造は、等価な階層的な符号化に置き換えることができ、各段階での符号化はスカラー量子化によって表現できるからである。他にも shape-gain ベクトル量子化 [82] や変換パラメータをそれぞれスカラー量子化する伝統的な変換符号化の手法 [83] [81] も product code の一種である。これらの手法の存在は、product code が最適ではないにしろ実用的には十分な圧縮性能を与えることのできる符号化方法であることを示している。

以下、5.2 節では多次元 companding 関数をパラメタライズし、そのパラメータを情報源のサンプルから適応的に学習させることによりパラメータ族の中で最適な符号化を構成す

る方法を提案する。この方法では、一般的な定常分布に従う情報源に対してパラメタライズされた多次元 companding 関数のクラスの中から歪みを最小化する多次元 companding 関数を求めることができる。ただし、ブロック内の情報は定常情報源から、かつブロック間で互いに独立な標本であると仮定する。提案する符号化方法では量子化器に一様量子化器を用いており、その量子化数を増減することで容易にレート制御が行えることもメリットで、5.3 節では歪み基準を二乗誤差基準としたときのレートの増減法について提案する。5.4 節で本手法の有効性を確かめる簡単なシミュレーションを行い、5.5 節でまとめを述べる。

5.2. product code モデル

5 節全体での表記法を示す。$\mathbf{x} = [x_1, \cdots, x_n]^T \in R^n$ を確率密度分布 $p(\mathbf{x})$ に従う n 次元ベクトルの情報源とし、$\hat{\mathbf{x}} \in \{\hat{\mathbf{x}}\} \equiv \left\{ \hat{\mathbf{x}}^1, \cdots, \hat{\mathbf{x}}^M \middle| \hat{\mathbf{x}}^k = [\hat{x}, \cdots, \hat{x}_n]^T \in R^n, k = 1, \cdots, M \right\}$ を reproduction とする。ただし、$[\cdot]^T$ はベクトルの転置を表す。Reproduction $\hat{\mathbf{x}}$ は M 通りの値をとるのでこれを符号化するのに要する符号長は $\log_2 M$bit である。Product code で可変長符号化、すなわち量子化数を入力情報に応じて選択することは可能であるが、複雑となるのでここでは固定長符号化のみを考える。また reproduction $\hat{\mathbf{x}}$ は、情報源 $\mathbf{x}$ から符号化関数 $\hat{\mathbf{x}} = \rho(\mathbf{x})$ によって得られるとする。$\mathbf{x}$, $\hat{\mathbf{x}}$ の属する空間をまとめて入力空間と呼ぶ。情報源 $\mathbf{x}$ と reproduction $\hat{\mathbf{x}}$ の間の歪み測度は $d(\mathbf{x}, \hat{\mathbf{x}})$ で表し、$d(\mathbf{x}, \hat{\mathbf{x}}) \geq 0$ である。

5.2.1 product code の構成

特徴ベクトル $\mathbf{y} = [y_1, \cdots, y_m]^T$, $\hat{\mathbf{y}} = [\hat{y}_1, \cdots, \hat{y}_m]^T$ をそれぞれの要素の定義域が $(0, 1)$ である m 次元ベクトルとする。多次元 companding 関数は、特徴ベクトル $\mathbf{y}$、$\hat{\mathbf{y}}$ を用いて、compressor $\mathbf{y} = \psi(\mathbf{x}; \theta_\alpha)$ と expander $\hat{\mathbf{x}} = \phi(\hat{\mathbf{y}}; \theta_\beta)$ の組み合わせとして定式化される。また、compressor、expander それぞれの要素ごとの出力関数は $y_i = \psi_i(\mathbf{x}; \theta_\alpha)$、$\hat{x}_i = \phi_i(\hat{\mathbf{y}}; \theta_\beta)$ と表す。ただし、θ_α と θ_β はそれぞれ compressor と expander のパラメータであり、パラメータ依存性を強調する必要がない場合は、適宜省略する。特徴ベクトル $\mathbf{y}$ はランダム変数 $\mathbf{x}$ の関数なので、$\mathbf{y}$ もまたランダム変数となり、その確率密度を $p(\mathbf{y})$ で表す。すなわち compressor $\mathbf{y} = \psi(\mathbf{x})$ は、$R^n \to (0, 1)^m$ の写像、expander $\hat{\mathbf{x}} = \phi(\hat{\mathbf{y}})$ は、$(0, 1)^m \to R^n$ の写像である。関数 $\hat{\mathbf{y}} = \Gamma(\mathbf{y}) = [\Gamma_1(y_1), \cdots, \Gamma_m(y_m)]^T$ は $\mathbf{y}$ から $\hat{\mathbf{y}}$ への量

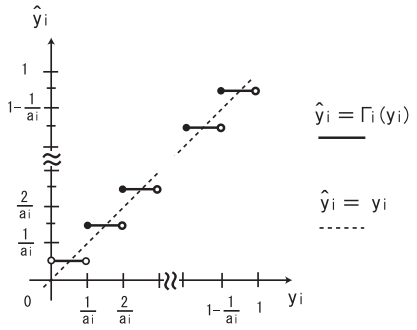


図 5.1 スカラー量子化関数 $\hat{y}_i = \Gamma_i(y_i)$。

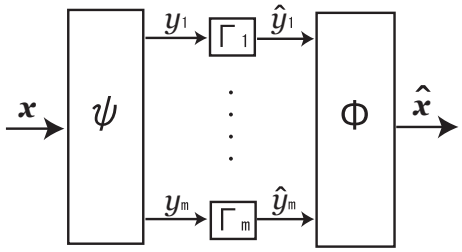


図 5.2 提案する product code の構成。

子化関数であり、m 個の区間 $(0,1)$ を互いに独立に一様量子化するスカラー量子化関数 $\Gamma(y_i) = \arg\min_{c \in D_i} |y_i - c|$ から成る。ただし、$D_i = \left\{ \frac{1}{2a_i}(2j-1) \,|\, j = 1, \cdots, a_i \right\}$ である。ここで、a_i は第 i 成分の量子化数 (正の整数) であり、$M = \prod_{i=1}^{m} a_i$ である。ただし、$\arg\min_{c \in D_i} |y_i - c|$ が複数存在するとき、すなわち $y_i = \frac{j}{a_i}$ のときは $\Gamma(y_i = \frac{j}{a_i}) = \frac{1}{2a_i}(2j+1)$ とする。$\hat{\mathbf{y}}^k = \left[\hat{y}_1^k, \cdots, \hat{y}_m^k \right]^T$ を k 番目の reproduction ベクトル $\hat{\mathbf{x}}^k$ と $\hat{\mathbf{x}}^k = \phi(\hat{\mathbf{y}}^k)$ のように対応する特徴ベクトルとすると、これは関数 $l_i(k) : \{k | k \in \{1, \cdots, M\}\} \to \{j | j \in \{1, \cdots, a_i\}\}$ を用いて、$\hat{\mathbf{y}}^k = \left[\frac{1}{2a_1}(2l_1(k)-1), \cdots, \frac{1}{2a_m}(2l_m(k)-1) \right]^T$ と書ける。

スカラー量子化関数 Γ_i を図 5.1 に図示する。以上を用いて、product code を行う符号化関数 ρ は $\rho(\mathbf{x}) = \phi(\Gamma(\psi(\mathbf{x})))$ と書ける。図 5.2 にこれらの companding 関数を用いた product code の構成を示す。

5.2.2 companding 関数の学習

最適な companding 関数を解析的に求めることは難しいが、companding 関数をパラメタライズしパラメータ族の中での局所最適解を求めることは、勾配法などの最適化手法により可能である。固定長符号化において符号長 $\log_2 M$bit が与えられたとき、歪みあり符号化の目的は次式のコスト関数の最小化となる。

$$\min_{\rho} E[d(\mathbf{x}, \rho(\mathbf{x}))] \tag{5.1}$$

ここで、E は $\mathbf{x}$ の分布 $p(\mathbf{x})$ での期待値を表す。パラメタライズされた我々の product code においては式 (5.1) は以下のようになる。

$$\min_{\theta_\alpha, \theta_\beta} E[d(\mathbf{x}, \phi(\Gamma(\psi(\mathbf{x}; \theta_\alpha)); \theta_\beta)] \tag{5.2}$$

この最適化問題には、勾配法やニュートン法などよく知られた方法の適用が考えられるが、そのためにはコスト関数がパラメータ $\theta_\alpha, \theta_\beta$ に対して適当階数、微分可能である必要がある。Compressor ψ, expander ϕ は微分可能であったとしても量子化関数 Γ は微分可能でない。そのため、パラメータ θ_β の学習には勾配法やニュートン法などの適用が可能であるのに対して、パラメータ θ_α の学習にはこれらの手法が適用できない。そこで、パラメータ θ_α の学習時には量子化関数 Γ を連続関数で近似することを考える。近似は次式のように単純な線形関数に置き換えることで行う。

$$\hat{\mathbf{y}} = \Gamma(\mathbf{y}) \approx \mathbf{y} \tag{5.3}$$

図 5.1 からも直感的に明らかなように、高量子化の極限 ($a_i \to \infty$, $i = 1, \cdots, m$) ではこの近似は正確なものとなる。つまり、高量子符号化時の近似とみなすことができる。

式 (5.2) には情報源 $\mathbf{x}$ の確率密度 $p(\mathbf{x})$ についての期待値計算が含まれているが、実際には情報源の確率分布は不明でその代わりに情報源からのサンプル $\mathbf{x}_1, \cdots, \mathbf{x}_N$ が得られる場合がほとんどである。各々のサンプル $\mathbf{x}_t$ が分布 $p(\mathbf{x})$ から独立に得られるとするとこれらのサンプルから期待値を $E[d(\mathbf{x}, \rho(\mathbf{x}))] \approx \frac{1}{N} \sum_{i=1}^{N} d(\mathbf{x}_i, \rho(\mathbf{x}_i))$ で概算し、このサンプル平均に基づきパラメータ $\theta_\alpha, \theta_\beta$ の最適化が可能となる。$T_N = \frac{1}{N} \sum_{i=1}^{N} d(\mathbf{x}_i, \rho(\mathbf{x}_i))$ は中心極限定理によって平均値 $E[T_N] = E[d(\mathbf{x}, \rho(\mathbf{x}))]$、分散 $\frac{V(T_1)}{N}$ の正規分布にほぼ従う。ここで、$V(T)$ はランダム変数 T の分散である。十分なサンプルが得られている場合 ($N \gg V(T_1)$)、分散 $\frac{V(T_1)}{N}$ が 0 に近くなり $E[d(\mathbf{x}, \rho(\mathbf{x}))] = \frac{1}{N} \sum_{i=1}^{N} d(\mathbf{x}_i, \rho(\mathbf{x}_i))$ と考えてよ

い。また分散 $\frac{V(T_1)}{N}$ は、与えられた情報源 $\mathbf{x}$ と符号化関数 $\rho(\mathbf{x})$ の下では近似の精度が主にサンプル数 N に依存し、情報源 $\mathbf{x}$ の次元 n によらないことを示していることにも注意する。ただし、厳密には分散 $V(T_1)$ が $\mathbf{x}$ の次元の増加に対して変化しないことを示す必要がある。例えば、一般的に広く用いられる加法的な歪み測度 $d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n}\sum_{i=1}^{n} d(x, \hat{x})$ を用いることを考え、さらにそれぞれの歪み $\xi_i \equiv d_i(x_i, \hat{x}_i)$ が互いに独立で同じ分布に従うランダム変数 ξ とすると、分散は $V(T_1) = V(\xi)$ となり $\mathbf{x}$ の次元に依存しない。

ベクトル量子化において、入力 1 次元当たりの符号長をレート R(bit) とすると、reproduction ベクトル集合 $\{\hat{\mathbf{x}}\}$ の要素数 M が $M = 2^{Rn}$ と入力次元 n の増大と共に指数的に増大する。最近傍ベクトルの探索 $\arg\min_{\hat{\mathbf{x}}^k} d(\mathbf{x}, \hat{\mathbf{x}}^k)$ はベクトル量子化の最適性のための必要条件である最近傍条件を満足させるために必要であるが、この探索には $\{\hat{\mathbf{x}}\}$ の要素数 $M = 2^{Rn}$ に比例した計算コストがかかる。これに対して我々の符号化では、最近傍ベクトルを探索する代わりに compressor $\psi(\mathbf{x}; \theta_\alpha)$ と量子化関数 Γ によって、$\mathbf{x}$ に対する reproduction $\hat{\mathbf{x}}$ が決められる。仮に compressor $\psi(\mathbf{x}; \theta_\alpha)$ が要する計算量がパラメータ θ_α に依存せず c_{com} と書けるなら、関数 $\Gamma(\psi(\mathbf{x}; \theta_\alpha))$ に要する計算量 c は c_{com} と量子化 $\Gamma(\cdot)$ に必要な計算量 $\log_2 M$ との合計 $c = c_{com} + \log_2 M$ になる。$M = 2^{Rn}$ のとき $c = c_{com} + Rn$ である。一方で、compressor $\psi(\mathbf{x}; \theta_\alpha)$ に必要な計算量 c_{com} は compressor のモデルに依存する。表現能力の高い関数族のためにはモデルを複雑にする必要があるが、その代わり計算量は増大するので計算量はモデルの複雑さとのトレードオフで決められる。例えば、一般の線形変換による変換符号化の場合、$c_{com} = nm$ となる。

パラメータ θ_α と θ_β は、それぞれ一方を固定してもう一方を最適化するような繰り返し最適化法により求めることができる。パラメータ θ_β を固定してパラメータ θ_α を最適化すると、パラメータ θ_α は式 (5.2) と式 (5.3) の近似より、

$$\theta_\alpha = \arg\min_{\theta_\alpha} E\left[d(\mathbf{x}, \phi(\psi(\mathbf{x}; \theta_\alpha)))\right] \tag{5.4}$$

となるように決められる。一般に歪み測度 $d(\mathbf{x}_1, \mathbf{x}_2)$ は $\mathbf{x}_1$ と $\mathbf{x}_2$ が一致したときのみ 0 をとり、それ以外で正の値をとるので $\mathbf{x} = \phi(\psi(\mathbf{x}; \theta_\alpha))$ となるとき最小の歪みを与える。すなわち、もし expander ϕ が逆関数 ϕ^{-1} をもつならば $\psi = \phi^{-1}$ となるようにパラメータ θ_α を決めればよい。この場合、パラメータ θ_β さえ決まれば常にパラメータ θ_α は決定されるので θ_α は探索する必要がなくなり、探索時間削減や局所最適解の回避にメリットとなる。

5.2.3 平均歪みの評価

ここでは歪み測度 d を加法的な二乗誤差として、そのときの平均歪みを議論する。すなわち

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \tag{5.5}$$

となるような歪み測度をとる。

$$\begin{aligned} \mathbf{y} &= \psi(\mathbf{x}), \\ \mathbf{x} &= \phi(\mathbf{y}) \end{aligned} \tag{5.6}$$

が成り立つとき、つまり ψ が逆関数 ψ^{-1} をもち、$\phi = \psi^{-1}$ として $\mathbf{x}$, $\mathbf{y}$ が一対一対応するとき、式 (5.6) を式 (5.5) に代入して、

$$E[d(\mathbf{x}, \hat{\mathbf{x}})] = \frac{1}{n} E\left[\sum_{i=1}^{n} \left(\phi_i(\mathbf{y}) - \phi_i(\hat{\mathbf{y}}) \right)^2 \right] \tag{5.7}$$

となる。

通常、特徴ベクトル $\mathbf{y}$ は情報源 $\mathbf{x}$ の圧縮表現となるのでその次元 m は情報源の次元 n より小さくなり、その場合 $\mathbf{x}$ と $\mathbf{y}$ が一対一対応するような逆関数 ϕ^{-1} はしばしば存在せず、式 (5.7) のような表現はできない。しかし、特徴空間は情報源 $\mathbf{x}$ ができるだけ表現できるように選ばれるので、よく出現する $\mathbf{x}$ の領域については、$\mathbf{x} = \phi(\mathbf{y})$ となるある $\mathbf{y} \in (0,1)^m$ が存在すると考えてよい。すなわち領域 D を $D = \{\mathbf{x} | \mathbf{x} = \phi(\mathbf{y}), \mathbf{y} \in (0,1)^m\}$, 全空間を R^n として D^c を D の補空間とした時に、$\int_{D^c} p(\mathbf{x}) d\mathbf{x} \approx 0$ と近似できるなら、$\mathbf{x} \in D^c$ を無視して式 (5.7) が成り立つと考えてよい。

Expander ϕ が $\mathbf{y}$ について 2 階微分可能のとき、次式のように $\phi_i(\hat{\mathbf{y}})$ を点 $\mathbf{y}$ の周りでテイラー展開できる。

$$\phi_i(\hat{\mathbf{y}}) = \phi_i(\mathbf{y}) + \sum_{j=1}^{m} \phi'^j_i(\mathbf{y})(\hat{y}_j - y_j) + \Delta^2 \tag{5.8}$$

ここで、$\phi'^j_i(\mathbf{y}) = \frac{\partial \phi_i(\mathbf{y})}{\partial y_j}$ であり、Δ^2 は $(\hat{y}_j - y_j)$ について 2 次の項を指す。$\mathbf{y}$ の各要素 y_i が区間 $L_i^k \equiv \left[\hat{y}_i^k - \frac{1}{2a_i}, \hat{y}_i^k + \frac{1}{2a_i} \right)$, $k = 1, \cdots, M$ となる m 次元立方格子 C_k 内で $\phi_i(\mathbf{y})$ が線形近似できるなら Δ^2 の項は無視できる。

今、各立方格子 C_k 内で expander $\phi_i(\mathbf{y})$ が線形近似できるほど十分滑らかで Δ^2 の項

が無視できるとして、式 (5.8) を式 (5.7) に代入すると、

$$\begin{aligned} E[d(\mathbf{x},\hat{\mathbf{x}})] &\approx \frac{1}{n}E\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{m}\phi'^{j}_{i}(\mathbf{y})(\hat{y}_j - y_j)\right)^2\right] \\ &= \frac{1}{n}\sum_{k=1}^{M}\int_{\mathbf{y}\in C_k} p(\mathbf{y})\sum_{i=1}^{n}\left(\sum_{j=1}^{m}\phi'^{j}_{i}(\mathbf{y})(\hat{y}^k_j - y_j)\right)^2 d\mathbf{y} \end{aligned} \tag{5.9}$$

となる。さらに各立方格子 C_k 内で分布 $p(\mathbf{y})$ と $\phi'^{j}_{i}(\mathbf{y})$ が一定とみなせるとし、それぞれ p_k と $\phi'^{j}_{i\,k}$ で表すと式 (5.9) は以下のようになる。

$$\begin{aligned} E[d(\mathbf{x},\hat{\mathbf{x}})] &\approx \frac{1}{n}\sum_{k=1}^{M}\int_{\mathbf{y}\in C_k} p_k \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\phi'^{j}_{i\,k}\right)^2(\hat{y}^k_j - y_j)^2 d\mathbf{y} \\ &= \frac{1}{n}\sum_{j=1}^{m}\sum_{k=1}^{M}\int_{\mathbf{y}\in C_k} p_k G^k_j (\hat{y}^k_j - y_j)^2 d\mathbf{y} \\ &= \frac{1}{n}\sum_{j=1}^{m}\sum_{k=1}^{M}\left\{\int_{\bar{\mathbf{y}}^j\in\bar{C}^j_k} p_k G^k_j d\bar{\mathbf{y}}^j \int_{y_j\in L^k_j}(\hat{y}^k_j - y_j)^2 dy_j\right\} \end{aligned}$$

$a_j\int_{y_j\in L^k_j} dy_j = 1$ を各項に掛けて

$$\begin{aligned} E[d(\mathbf{x},\hat{\mathbf{x}})] &\approx \frac{1}{n}\sum_{j=1}^{m}\frac{1}{12a_j^2}\sum_{k=1}^{M}\int_{\mathbf{y}\in C_k} p_k G^k_j d\mathbf{y} \\ &= \frac{1}{n}\sum_{j=1}^{m}\frac{1}{12a_j^2}E\left[G_j(\mathbf{y})\right] \end{aligned} \tag{5.10}$$

ここで、$G^k_j = \sum_{i=1}^{n}\left(\phi'^{j}_{i\,k}\right)^2$, $G_j(\mathbf{y}) = \sum_{i=1}^{n}\left(\phi'^{j}_{i}(\mathbf{y})\right)^2$ であり、$\bar{\mathbf{y}}^j$ と $\bar{C}^j_k$ はそれぞれ特徴ベクトル $\mathbf{y}$ から y_j 成分を除いたベクトルと立方格子 C_k から y_j 軸を除いた部分格子である。式 (5.10) には、$p(\mathbf{x})$ についての期待値計算が含まれているが、パラメータ学習の場合と同様にサンプル平均により推定することができる。

5.2.4 ビット割り当て

特徴ベクトル $\mathbf{y}$ の各要素 y_i を一様量子化するための量子化レベル a_i の決定について考える。量子化レベルを変えるとき、符号長 $\sum_{i=1}^{m}\log_2 a_i$ もそれに応じて変化する。こ

のとき、符号長の制約 $\sum_{i=1}^{m} \log_2 a_i \leq \log_2 M$ の下での最適化より、符号長の制約を取り払ったレート歪みの指標 $RD(\rho)$ での最適化のほうが扱いやすい。

$$RD(\rho) = \sum_{i=1}^{m} \log_2 a_i + \gamma E[d(\mathbf{x}, \rho(\mathbf{x}))] \tag{5.11}$$

ここで γ はあらかじめ決められたある正の実数である。このコスト関数 $RD(\rho)$ を小さくする符号器 ρ ほど良い符号器ということにする。いったん量子化レベルが決まれば、$RD(\rho)$ は式 (5.1) と同じコスト関数となり、ここまでの最適化の議論などはそのまま適用できる。

式 (5.11) に関して大域的に最適な量子化レベルを求めるためには、式 (5.10) による平均歪みの推定値を使ったとしても、最適な companding 関数が何であるかをあらかじめ知る必要があり、容易でない。しかし局所的な量子化レベルの増減では、その量子化レベルでの最適な companding 関数が局所的増減ではほとんど変わらないと仮定できるので、平均歪みの変化は式 (5.10) に基づいて推定することができる。ここでは局所的増減として、ある要素 y_i の量子化レベル a_i を 1 つ減少させる、もしくはある要素 y_j の量子化レベル a_j を 1 つ増加させることを考える。これらの量子化レベルの局所的増減によって全体のコスト $RD(\rho)$ を減らすことができれば、レート歪みの観点からより良い符号器を構成できたとみなすことにする。

式 (5.10) より、ある要素 y_i の量子化レベル a_i を 1 つ減らしたときのコスト関数 $RD(\rho)$ の変化 Δ_{-i} は、

$$\Delta_{-i} = \log_2 \left(1 - \frac{1}{a_i}\right) + \gamma \frac{E\left[G_i(\mathbf{y})\right]}{12n} \frac{2a_i - 1}{(a_i - 1)^2 a_i^2}$$

となる。同様に、ある要素 y_j の量子化レベル a_j を 1 つ増やしたときのコスト関数 $RD(\rho)$ の変化 Δ_{+j} は、

$$\Delta_{+j} = \log_2 \left(1 + \frac{1}{a_j}\right) - \gamma \frac{E\left[G_j(\mathbf{y})\right]}{12n} \frac{2a_j + 1}{(a_j + 1)^2 a_j^2}$$

となる。

これら $\{\Delta_{-i}, \Delta_{+j}\}$ のうち最も小さい変化分 Δ を求め、$\Delta < 0$ ならば対応する量子化レベルの変化を行うことで、より良い符号器 ρ が構成できる。この手続きを繰り返すことで、符号器の改良を行う。

5.3. シミュレーション

単純な companding 関数を用いたときの product code の性能を、KL 変換を用いた変換符号化の性能と比較した。図 5.3 に示すように情報源は二次元 $\mathbf{x} = [x_1, x_2]^T$ で、二次元の一様分布 $\mathbf{z} = [z_1, z_2]^T$ を 2×2 行列 $\mathbf{K}$ によって線形変換 $\mathbf{x} = \mathbf{K}\mathbf{z}$ したものを用いた。ここで用いた companding 関数は、

$$\begin{aligned} \psi(\mathbf{x}; A) &= f(A^{-1}\mathbf{x}) \\ \phi(\hat{\mathbf{y}}; A) &= Ag(\hat{\mathbf{y}}) \end{aligned}$$

である。ただし、関数 $f(\mathbf{x})$ は二次元ベクトル関数であり、そのスカラー関数 $f_i(\mathbf{x})$ $(i = 1, 2)$ は

$$f_i(x_i) = \left\{ \begin{array}{cc} 1 & (1 < x_i) \\ \frac{1}{2}x_i + \frac{1}{2} & (-1 \leq x_i \leq 1) \\ 0 & (x_i < -1) \end{array} \right\}$$

である。関数 $f(\mathbf{x})$ の値域が $[0, 1]$ となっているが、5.2 節で compressor を $R^n \rightarrow (0, 1)^m$ のように値域が $(0, 1)^m$ の関数として定義したことは本質ではなく、$[0, 1]^m$ となっても同様の議論が可能である。このとき expander は $[0, 1]^m \rightarrow R^n$ の関数となる。関数 $g(\mathbf{x})$ もまた二次元のベクトル関数であり、そのスカラー関数 $g_i(\mathbf{x})$ $(i = 1, 2)$ は

$$g_i(x_i) = 2x_i - 1, (0 \leq x_i \leq 1)$$

である。行列 A は 2×2 の正則な行列である。A^{-1} は A の逆行列を表す。ここでは expander ϕ が逆関数 ϕ^{-1} をもつので、5.2.2 節の最後で述べたように compressor ψ のパラメータは、expander ϕ の逆関数 ϕ^{-1} となるように固定している。

この companding 関数においては行列 A がパラメータとなっており、$A^* = \arg\min_A \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i - A\mathbf{z}_i)^2$ となる A^* を求める。ここで、$\mathbf{x}_i$ は情報源 $\mathbf{x}$ からの i 番目のサンプルであり、$\mathbf{z}_i$ は $\mathbf{z}_i = g\left(\Gamma\left(f(A^{-1}\mathbf{x}_i)\right)\right)$ である。A^* は行列 $X = [\ \mathbf{x}_1,\ \cdots\ , \mathbf{x}_N\ ]$ と $Z = [\ \mathbf{z}_1,\ \cdots\ , \mathbf{z}_N\ ]$ を用いて $A^* = XZ^\dagger$ とかける。ただし、$Z^\dagger$ は行列 Z の最小二乗型擬逆行列である。$\mathbf{z}_i$ はパラメータ A と無関係な定数としてパラメータ A を最適化したが、実際には相関があるので収束値付近で振動を引き起こしやすい。この影響を軽減するため、実際には $A_{t+1} = \eta * A_t + (1 - \eta)A^*$ のように慣性項を加えて更新した。ただし、η は 0.1 程度のスカラー値で A_t は t 番目に更新した際のパラメータ A を表す。

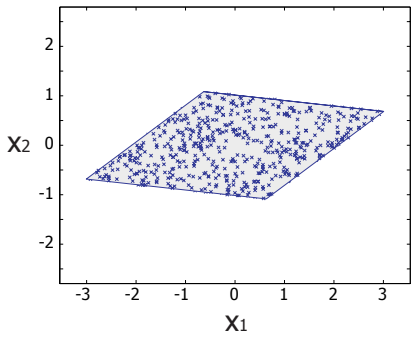


図 5.3　二次元情報源の分布 (標本数 1000 点)。

KL 変換を用いた変換符号化では、KL 変換 $\mathbf{y} = \mathbf{K}\mathbf{x}$ を行った後、変換後の特徴ベクトル $\mathbf{y}$ に対してスカラー量子化を行った。前述のようにスカラー量子化を高量子符号化する極限では、companding 関数 $F_i(y_i) = \frac{1}{Z}\int_{-\infty}^{y_i} p_i(z)^{1/3}dz$ によって最適な量子化が可能である。ただし、$Z = \int_{-\infty}^{+\infty} p_i(z)^{1/3}dz$, $p_i(z)$ は y_i を標準偏差 σ_i で正規化した $z = \frac{y_i}{\sigma_i}$ の確率密度である。符号長 $B = \log_2 M$ ビット以内で符号化を行う場合、$\sum_{i=1}^{n} b_i \leq B$, $b_i \geq 0$ の制約の下で第 i 番目のスカラー量子化を b_i ビット $(i = 1, \cdots, n)$ で行うことができる。このとき最適なビット割り当ては $b_i = \bar{b} + \frac{1}{2}\log_2 \frac{\sigma_i^2}{\rho^2} + \frac{1}{2}\log_2 \frac{h_i}{H}$ で与えられる。ただし、$\bar{b} = \frac{B}{n}$, $\rho^2 = \left(\prod_{i=1}^{n} \sigma_i^2\right)^{1/n}$, $h_i = \frac{1}{12}\left\{\int_{-\infty}^{+\infty} p_i(z)^{1/3}dz\right\}^3$, $H = \left(\prod_{i=1}^{n} h_i\right)^{1/n}$ である [84]。このビット割り当て解は必ずしも正の整数の量子化レベルとはならないが、そのときは近い整数の値を候補とし、その候補から最も歪みが小さくなる整数を選ぶ。Product code の最適化では、情報源の確率密度の知識を必要とせず、情報源からのサンプルによって最適化可能であるのに対し、上記の変換符号化では情報源の確率密度の知識を必要としていることに注意する。

図 5.4 に product code による符号化と KL 変換を用いた変換符号化の結果を示す。比較のために、どちらもある符号長の制約のもとでの平均誤差を求めたが 3.2 節で述べたように product code による符号化では式 (5.11) で表されるレート歪み指標 RD に基づいて最適化を行っている。レート歪み指標 RD の γ は、符号長が 4 ビット以内,6 ビット以内の符号化のとき、それぞれ $\gamma = 25$,$\gamma = 100$ に設定している。どちらの指標で最適化を行っても歪み・レート関数 $D^* = D(R^*)$ 上の値 (R^*, D^*) を与える。

情報源 $\mathbf{x}$ の定義域は図 5.3 に示す様に、平行四辺形内であり、reproduction ベクトル

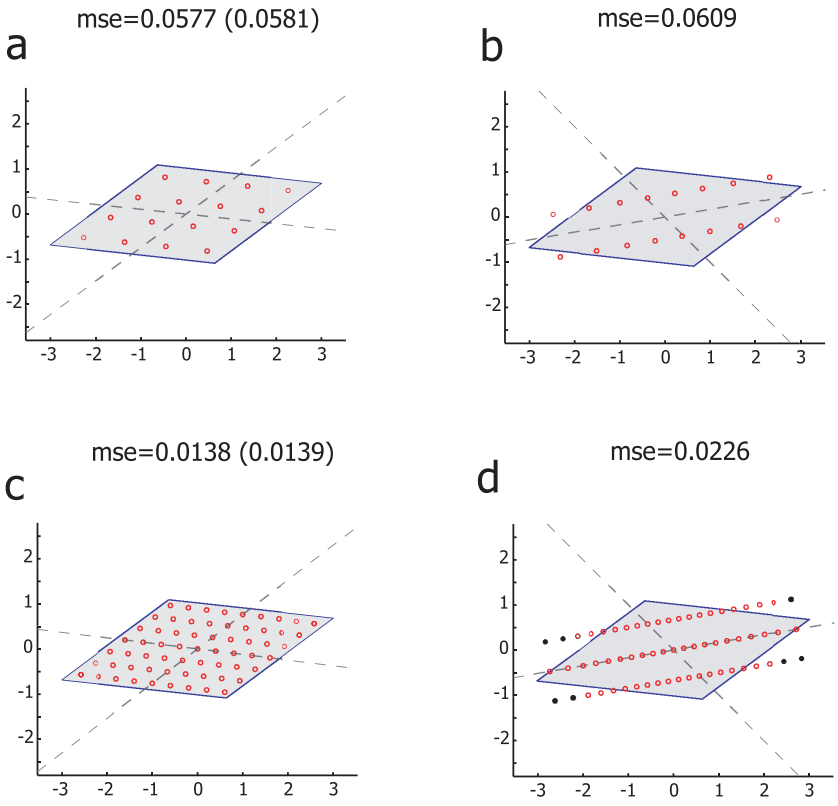


図 5.4 Product code と KL 変換を用いた変換符号化の結果。(a)：符号長が 4 ビット以内での product code による符号化。(b)：符号長が 4 ビット以内での KL 変換を用いた変換符号化による符号化。(c)：符号長が 6 ビット以内での product code による符号化。(d)：符号長が 6 ビット以内での KL 変換を用いた変換符号化による符号化。

の集合 $\{\hat{\mathbf{x}}\}$ が丸印で示されている。黒塗りの丸印は、reproduction ベクトルのうち情報源 $\mathbf{x}$ の符号化に使われない不要な reproduction ベクトル、すなわち空乏セルを表す。図 5.4a,b は共に符号長が 4 ビット以内での符号化結果であり、図 5.4c,d は共に符号長が 6 ビット以内での符号化結果である。また図 5.4a,c が本論文で提案した product code による符号化結果で、図中の直線が特徴ベクトル y_1, y_2 に沿う軸を表す。図 5.4b,d が KL 変換を用いた変換符号化の結果で、図中の直線が主軸を表している。それぞれの図上に、その符号化を行ったときの平均二乗誤差 (mse) を示す。また、図 5.4a,c の図上に式 (5.10) を用いて推定した平均二乗誤差を括弧書きで示す。Product code で学習された線形変換と KL 変換の結果の違いが図中の直線 (特徴ベクトル) の向きからわかる。Product code に

おける線形変換では、特徴ベクトル y_1, y_2 が互いに独立となるような変換を行うことですべての reproduction ベクトルが無駄なく使われるようにしているが、KL 変換では変換後のベクトルが互いに無相関にはなるものの独立とはならない。低ビット符号化時には、どちらの符号化によっても reproduction ベクトルが情報源の定義域内に収まり、平均二乗誤差にそれほど差がでなかったが、高ビット符号化時には KL 変換を用いた変換符号化に空乏セルが現われることで、結果的に平均二乗誤差が大きく (product code の約 1.6 倍に) なる。

同じデータに対して、GLA で最適化したベクトル量子化による平均二乗誤差は、4 ビット以内、6 ビット以内のそれぞれの符号化で、0.040、0.010 となった。もしも、product code で求めた reproduction ベクトルに対してベクトル量子化で用いたのと同じ最近傍法で符号化を行うと、平均二乗誤差は 4 ビット以内、6 ビット以内のそれぞれの符号化で、0.041、0.010 となり、ベクトル量子化によるものとほとんど差はなくなった。これは、product code とベクトル量子化の性能の違いが主に計算コストの高い最近傍法を用いるかどうかに起因していることを示している。すなわち、最適な reproduction ベクトルは product code によって十分良く表現できている。

5.4. まとめ

本章では、多次元 companding 関数の学習による product code の設計法を提案した。線形変換を companding 関数に採用したときの product code の性能を、KL 変換による変換符号化と比較し、その有効性を確かめた。KL 変換を用いた変換符号化は、情報源がガウス分布であるという仮定のもとでは、あらゆる直交線形変換による変換符号化の中で最適となるが、シミュレーションに用いた情報源はガウス分布ではない。今回の結果は、情報源が非ガウス分布である場合において、companding 関数の学習によって優れた変換符号器が得られることを示している。また、product code で符号化を行ったときの平均歪みの推定も行った。実験に用いたデータでは companding 関数が線形であったことや、情報源が局所的に一様に分布していたため、式 (5.10) の導出のための仮定が満たされることになり、平均二乗誤差が非常によく推定できた。Product Code の特徴は、量子化が特徴ベクトル $\mathbf{y}$ のそれぞれの要素 y_i に独立に行われることにある。量子化によって平均歪みは決定されるが、符号長を決定するものではない。ここでは、単純に符号長は $\sum_{i=1}^{m} \log_2 a_i$ と仮定している。$\hat{\mathbf{y}}_j$ の生じる確率が分布 $p(\hat{\mathbf{y}}_j)$ で与えられたとき、平均符

号長を最小にするためには $\hat{\mathbf{y}}_j$ に割り当てる符号長は $\log p(\hat{\mathbf{y}}_j)$ にするべきである。ここでは、分布が与えられている代わりに符号長を仮定している。そして、その符号長はどの $\hat{\mathbf{y}}_j$ に対しても同じ符号長である。そのため、もし符号長が可変だったとしてもそれ以上、符号長を小さくできないという意味で無駄がない符号化は分布 $p(\hat{\mathbf{y}})$ が一様分布のときであるといえる。また、符号長を $\sum_{i=1}^{m} \log_2 a_i$ と仮定しない場合でも、Product Code という符号化をとる限り、量子化は特徴ベクトル $\mathbf{y}$ のそれぞれの要素 y_i に独立に行われる。このときの無駄のない符号化は分布 $p(\hat{\mathbf{y}})$ が独立なときである。そのため、product code による効率的な圧縮表現は、特徴ベクトル $\mathbf{y}$ が独立な表現となりやすいと考えられる。ただし、この議論では歪みの影響を考えていない。直交変換以外の線形変換を考えるとき変換後の変数で測られる歪みは変換前の変数で測られる歪みと異なってきてしまうので、どのような場合に独立な表現となりやすいかを定量的に知ることは難しい。また、スパースコーディングは期待対数尤度 Q の逐次最大化、

$$\begin{aligned} Q &= \int p(\mathbf{s}|\mathbf{x}) \log p(\mathbf{x}, \mathbf{s}) d\mathbf{s}, \\ &= \int p(\mathbf{s}|\mathbf{x}) \log p(\mathbf{s}|\mathbf{x}) d\mathbf{s} + \int p(\mathbf{s}|\mathbf{x}) \log p(\mathbf{s}) d\mathbf{s} \\ &\approx |\mathbf{x} - A\hat{\mathbf{s}}|^2 + R(\hat{\mathbf{s}}) \end{aligned}$$

と定式化されるが、第 1 項は歪みによるペナルティ、第 2 項は符号長に対するペナルティを表したものと解釈することができる。第 2 項は、スパースな事前分布をおいているため $\hat{\mathbf{s}}$ は 0 に近い値をとりやすい。このようにスパースコーディングなどの脳の情報表現モデルと圧縮符号の生成が関連していることがわかる。

第6章　結言

本論文では、まず、エントロピー最大化に基づくモデルを音声や器楽音を含む音楽を学習データとして学習を行い、学習後のモデルによって virtual pitch とマスキングの 2 つの現象をシミュレーションを行った。このモデルは、2 つの階層的な構造をもち、生物学的にそれぞれ個々の有毛細胞の非線形な入出力応答とレイヤー間の双方向のシナプス結合に対応させて考えることができた。しかしながら、2 つの階層に分離させてそれぞれ独立にエントロピー最大化を行っているため、一貫したエントロピー最大化を行っている保証がなかった。また、積分計算を事後確率が MAP 値でデルタ関数近似できるという近似にすることで、生物学的に妥当なヘブ学習則を導出することができたが、近似精度は粗かった。このモデルによって virtual pitch を引き起こすと見られるニューロンの活動が模擬されたが、それは一部のニューロンにおいて観察されたのみであった。

そこで、4 章でより理論的に一貫したモデルを提案した。2 段階に分離していたエントロピー最大化モデルは、非線形ノイズつき ICA モデルとして定式化しなおされた。ノイズが含まれたとき、決定論的な変換によってエントロピーを最大化するという決定論的エントロピー最大化の前提が崩れてしまう。そこで、決定論的な変換に代わって確率的にサンプリングした値のエントロピーを最大化するという確率論的エントロピー最大化の問題に一般化した。確率論的エントロピー最大化問題は、最尤推定による分布推定に帰着することができるので、隠れ変数を含むモデルの最尤推定を EM アルゴリズムで行った。ここでも積分の計算の問題が現われるが、積分の計算は事後分布をガウス近似で置き換えるラプラス近似によって行った。この近似計算によって得られる非線形ノイズ付き ICA の学習則は、これまでの低ノイズ近似に基づくノイズ付き線形 ICA やノイズなし線形 ICA を含む一般的な形となっていた。3 章で提案したノイズ付き ICA とは、ガウス近似している部分をデルタ関数近似で置き換えると一致する。一方、3 章のノイズ付き ICA は、Olshausen ら [11] や Rao ら [12] のスパースコーディングに基づくモデルと同じ形式であったが、これらのノイズ付き ICA の手法はノイズなし ICA を含むような一般的形となっていなかった。すなわち、従来の手法はノイズなし ICA を一般化したノイズあ

りICAの問題として定式化されていたが、積分の近似計算において事後確率をデルタ関数により近似すると、得られる学習則はノイズなしICAを一般化するものとはなっていなかった。一方、4章で述べた新しい手法では、事後確率をガウス近似することで、得られる学習則がノイズなし、ノイズありICAを共に含む形となった。このことから、従来、スパースコーディングに基づくモデルは近似精度の粗さのため一般性を失っていたこと、および新しく導入されたガウス近似の有効性が結論付けられた。また、尤度計算に現われる積分の近似にガウス近似を用いるとノイズなしICAを含む学習則の導出はできていなかったが、これはガウス近似に含まれるMAP値がパラメータに依存するためではないかと考えられた。シミュレーション結果からは、また、音高知覚には非線形性が本質的に重要であることが確かめられた。モデルで得たvirtual pitchの存在領域と実験的に得られたvirtual pitchの存在領域とを比較することでモデルの妥当性が示された。

5章では、理論的に冗長性圧縮の原理がproduct codeという制約をもった構造の歪みあり符号と関係していることを示した。推定されている脳全体のニューロンの数から記憶容量を推定したところ、その記憶容量が現代の計算機と比べて小さいことが推測される。記憶容量に大きな制限がかかっていることからもレート歪み理論の基準が初期知覚系の情報表現を説明するのに妥当な基準であると考えられる。Product codeはその制約のために、ほとんどの場合、最適な歪みあり符号を表現することができないが、実用的には様々な形で使われており、最適でないにしろ十分な表現能力を持っていることが推測される。本論文では、パラメタライズされたcompanding関数を最適化することによって、product codeを最適化することを提案した。また、歪み基準が二乗誤差基準であるときのビット割り当ての最適化法も示した。提案したproduct codeの性能の評価のため、簡単なシミュレーションを行った。データの分布がガウス分布でありかつ量子化数が大きいとき線形変換では最適な圧縮符号を与えるため、一般的にKL変換が良いとされているが、シミュレーション実験では提案手法がKL変換を超える性能を示すことができた。しかしこの実験で仮定した線形なモデルでは、reproductionベクトルを規則的にしか配置することができないため，特に情報源が非一様に分布した場合に符号化効率が悪くなることが予想される。今後，情報源分布の非一様性に対応できるようにcompanding関数を非線形化し，実際の自然画像や音声データに対する符号化能力について調べていきたいと思っている。

これまで画像圧縮や音声圧縮の分野でフーリエ変換やウェーブレット変換がよく使われているが、これらの変換のメリットの一つに計算量が少ないことが挙げられる。$n \times n$

の行列による線形変換は $O(n^2)$ の計算量が必要であるが、フーリエ変換やウェーブレット変換においては $O(n \log n)$ の計算量での計算方法が確立されている。ベクトル量子化に比較すると大きな差ではないが、計算量の問題は実用上大きい。しかし、線形変換による計算は並列計算が可能であれば計算量は n に依存しない量にまで落とせ、大幅な高速化が可能である。脳は、記憶容量は多く見積もっても 12 ギガバイト程度とノートパソコンにも劣るほどの容量でしかないが、並列計算の能力は、市販のパソコンでは到底、太刀打ちできない。効率的な符号化のためには、おそらく脳が用いているように、複雑な計算であっても、並列的な回路によって計算量の問題を克服するという方向性が妥当と考えられる。

謝辞

本論文をまとめるに当たって多くの人のご支援とご教授を承りました。それらについて心から感謝したいと思います。石井先生は、私を信頼してくださり、必要なときに叱咤激励をしてくださいました。この研究を進めることができましたのは、ひとえに石井先生のご指導のおかげであります。この研究室でなければ今の研究はなかったものと思います。謹んで深謝申し上げます。柴田先生は、日ごろから研究の姿勢や研究に関する有益な情報を教えていただきました。ありがとうございました。作村先生は、快適な計算機環境を整えてくださりました。ありがとうございました。大羽先生からは、研究全般に関して有益な議論を得ることできました。ありがとうございました。また、鹿野先生、猿渡先生からは聴覚研究の立場から、関先生は主に符号研究の立場から重要な問題提起を頂きました。それら有益なアドバイスとともに論文の審査を引き受けてくださったことに深く感謝します。

参考文献

[1] J. McCarthy and P. Hayes: “Some philosophical problems from the standpoint of artificial intelligence”, Machine Intelligence, **4**, pp. 463–502 (1969).

[2] D. C. Dennett: “Cognitive Wheels: The Frame Problem of AI”, Cambridge University Press (1984).

[3] D. H. Hubel and T. N. Wiesel: “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”, J. Physiol., **160**, pp. 106–154 (1962).

[4] H. K. Shinichi Maeda, Shinji Inagaki and W.-J. Song: “Separation of signal and noise from in vivo optical recording in guinea pigs using independent component analysis”, Neuroscience Letters, **302**, pp. 137–140 (2001).

[5] 前田新一, 十時慎一郎, 時岡良, 福西宏有：“大脳皮質聴覚野における周波数時空間応答特性の ica 適用による神経活動源の推定”, 第 14 回生体生理工学シンポジウム, 名古屋工大 (1999).

[6] 前田新一, 十時慎一郎, 時岡良, 福西宏有：“光計測データの ica 適用による皮質聴覚野トノトピィの独立神経活動源の推定”, 日本神経回路学会第 9 回全国大会講演論文集, 大阪 (1999).

[7] H. B. Barlow: “Sensory Communication”, MIT press (1961).

[8] R. Linsker: “Self-organization in a perceptual network”, Computer, **21(3)**, pp. 105–117 (1988).

[9] A. J. Bell and T. J. Sejnowski: “The ‘independent components’ of natural scenes are edge filters”, Vision Research, **37(23)**, pp. 3327–3338 (1997).

[10] 岡島健治, 今岡仁：“情報量最大化と生体視覚細胞の受容野”, 電子情報通信学会論文誌 A, **J83-A(6)**, pp. 620–628 (2000).

[11] B. A. Olshausen and D. J. Field: “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”, Nature, **381**, pp. 607–609 (1996).

[12] R. P. Rao and D. H. Ballard: “Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects”, Nature Neuroscience, **2(1)**, pp. 79–87 (1999).

[13] Y. Linde, A. Buzo and R. M. Gray: “An algorithm for vector quantizer design”, IEEE Trans. Comm., **COM-28**, pp. 84–95 (1980).

[14] A. S. Bregman: “Auditory Scene Analysis”, MIT Press (1990).

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

[16] L. P. Van Noorden: “Minimun differences of level and frequency for perceptual fission of tone sequences abab”, J. Acoust. Soc. Am., **61(4)**, pp. 1041–1045 (1977).

[17] A. S. Bregman and J. Campbell: “Primary auditory stream segregation and perception of order in rapid sequences of tones”, Journal of experimental psychology, **89(2)**, pp. 244–249 (1971).

[18] P. H. Lindsay and D. A. Norman: “Human information processing : an introduction to psychology”, Academic Press (1977).

[19] J. L. Flanagan: “Pitch discremination for synthetic vowels”, J. Acoust. Soc. Am., **30(5)**, pp. 435–442 (1958).

[20] A. Seebeck: “Beobachtungen über einige bedingungen der eutstehung von tönen”, Ann. Phys. Chem., **53**, pp. 417–436 (1841).

[21] A. Seebeck: “Uber die sirene”, Ann. Phys. Chem., **60**, pp. 449–481 (1843).

[22] R. J. Ritsma: “Existence region of the tonal residue. I”, J. Acoust. Soc. Am., **42(1)**, pp. 191–198 (1962).

[23] H. L. F. von Helmholtz: “Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik”, Brownschweig (1863).

[24] J. C. R. Licklider: “Periodicity pitch and place pitch”, J. Acoust. Soc. Am., **26**, p. 945(A) (1954).

[25] W. R. Thurlow and A. M. Small: “Pitch perception for certain periodic auditory stimuli”, J. Acoust. Soc. Am., **27**, pp. 132–137 (1955).

[26] E. de Boer: “Pitch of inharmonic signals”, Nature, **178**, pp. 535–536 (1956).

[27] J. F. Schouten, R. J. Ristma and B. L. Cardozo: “Pitch of the residue”, J. Acoust. Soc. Am., **34**, pp. 1418–1424 (1962).

[28] R. D. Patterson: “Physical variables determining residue pitch”, J. Acoust. Soc. Am., **53**, pp. 1565–1572 (1973).

[29] F. L. Wightman: “Pitch and stimulus fine structure”, J. Acoust. Soc. Am., **54**, pp. 397–406 (1973).

[30] 力丸裕：“視覚と聴覚”, pp. 129–179, 岩波講座認知科学, 岩波書店 (1994).

[31] F. L. Wightman: “The pattern-transformation model of pitch”, J. Acoust. Soc. Am., **54**, pp. 407–416 (1973).

[32] J. L. Goldstein: “An optimum processor theory for the central formation of the pitch of complex tones”, J. Acoust. Soc. Am., **54**, pp. 1496–1516 (1973).

[33] E. Terhardt: “Pitch, consonance, and harmony”, J. Acoust. Soc. Am., **55**, pp. 1061–1069 (1974).

[34] E. de Boer: “Psychophysics and Physiology of Hearing”, chapter Pitch theory unified, pp. 323–335, Academic Press (1977).

[35] 吉岡真：“3 つの音高知覚理論の数学的関連付けと一般化線形モデル”, J. UOEH (産業医科大学雑誌), **2(4)**, pp. 515–528 (1980).

[36] A. M. Noll: "Short-time spectrum and "cepstrum" techniques for vocal-pitch detection", J. Acoust. Soc. Am., **36(2)**, pp. 296–302 (1962).

[37] A. V. Oppenheim and R. W. Schafer: "Hormomorphic analysis of speech", IEEE. trans. Audio. Electroacoust., **AU-16(2)**, pp. 221–226 (1968).

[38] I. C. Whitefield and D. Purser: "Microelectrode study of the medial geniculate body in unanesthetized free-moving cats", Brain Behav. Evol., **6(1)**, pp. 311–328 (1972).

[39] N. M. Weinberger, J. H. Ashe, R. Metherate, T. M. Mckenna, D. M. Diamond and J. Bakin: "Returning auditory cortex by learning : A preliminary model of receptive field plasticity", Concepts in Neuroscience, **1**, pp. 91–132 (1990).

[40] S. C. E. Recanzone, G. H. and M. M. Merzenich: "Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys", J. Neurosci, **13(1)**, pp. 87–103 (1993).

[41] D. Robertson and D. R. Irvine: "Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness", J. Comp. Neurol., **282(3)**, pp. 456–471 (1989).

[42] C. Pantev, M. Hoke, B. Lutkenhoner and K. Lehnertz: "Tonotopic organization of the auditory cortex: pitch versus frequency representation", Science, **246**, pp. 486–488 (1989).

[43] H. Riquimaroux and T. Hashikawa: "Units in the primary auditory cortex of the Japanese monkey can demonstrate a conversion and place pitch in the central auditory system", J. de physique IV, **C5**, pp. 419–425 (1994).

[44] R. J. Zatorre: "Pitch perception of complex tones and human temporal-lobe function", J. Acoust. Soc. Am., **84(2)**, pp. 566–572 (1988).

[45] W. B. Davenport, Jr.: "A study of speech probability distribution", MIT Tech. Report, **148**, (1950).

[46] A. J. Bell and T. J. Sejnowski: "An information maximization approach to blind separation and blind deconvolution", Neural Computation, **7(6)**, pp. 1129–1159 (1995).

[47] A. J. Hudspeth and D. P. Corey: "Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli", Proc. Natl. Acad. Sci. U S A., **74(6)**, pp. 2407–2411 (1977).

[48] S. Laughlin: "A simple coding procedure enhances a neuron's information capacity", Zeitschrift fur Naturforschung. C, J. biosciences, **36(9-10)**, pp. 910–912 (1981).

[49] A. P. Dempster, N. M. Laird and D. B. Rubin: "Maximum likelihood from incomplete data via the em algorithm", J. Royal Statistical Society. Series B, StatistiCalMethodological, **39**, pp. 1–38 (1977).

[50] M. Sato and S. Ishii: "On-line em algorithm for the normalized gaussian network", Neural Computation, **12(2)**, pp. 407–432 (2000).

[51] A. Belouchrani and J. F. Cardoso: "Maximum likelihood source separation by the expectation-maximization technique: deterministic and stochastic implementation", Proceedings of NOLTA, pp. 49–53 (1995).

[52] E. Zwicker and H. Fastl: "Psychoacoustics", Springer-Verlag (1990).

[53] J. A. Costalupes, E. D. Young and D. J. Gibson: "Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat", J. neurophysiology, **51**, pp. 1326–1344 (1984).

[54] M. A. Cohen, S. Grossberg and L. L. Wyse: "A spectral network model of pitch perception", IEEE Trans. Auto. Contr. Sys., **98(2)**, pp. 862–879 (1995).

[55] E. F. Chang and M. M. Merzenich: "Environmental noise retards auditory cortical development", Science, **300**, pp. 498–502 (2003).

[56] J. O. Pickles and S. D. Comis: "Role of centrifugal pathways to cochlear nucleus in detection of signals in noise", J. neurophysiology, **36**, pp. 1131–1137 (1973).

[57] J. O. Pickles: "Role of centrifugal pathways to cochlear nucleus in determination of critical bandwidth", J. neurophysiology, **39**, pp. 394–400 (1976).

[58] S. G. Brown M. C. Kujawa and M. C. Liberman: "Single olivocochlear neurons in the guinea pig. II. response plasticity due to noise conditioning", J. neurophysiology, **79**, pp. 3088–3097 (1998).

[59] S.-I. Amari and J. F. Cardoso: "Blind source separation - semiparametric statistical approach", IEEE Trans. Signal Processing, **45(11)**, pp. 2692–2700 (1997).

[60] S.-I. Amari, T.-P. Chen and A. Cichocki: "Stability analysis of learning algorithms for blind source separation", Neural Networks, **10(8)**, pp. 1345–1351 (1997).

[61] J. F. Cardoso: "Infomax and maximum likelihood for source separation", IEEE Trans. Auto. Contr. Sys., **4(4)**, pp. 112–114 (1997).

[62] H. H. Yang and S.-I. Amari: "Adaptive on-line learning algorithms for blind separation - maximum entropy and minimum mutual information", Neural Computation, **9**, pp. 1457–1482 (1997).

[63] M. S. Lewicki and T. J. Sejnowski: "Learning overcomplete representations", Neural Computation, **12**, pp. 337–365 (2000).

[64] O. Bermond and J. F. Cardoso: "Approximate likelihood for noisy mixtures", ICA '99, Aussois, France, pp. 325–330 (1999).

[65] S.-I. Amari, A. Cichocki and H. H. Yang: "A new learning algorithm for blind signal separation", Advances in Neural Information Processing Systems, Vol. 8, MIT Press, pp. 752–763 (1996).

[66] S. Douglas, A. Cichocki and S. Amari: "Bias removal technique for blind source separation with noisy measurements", Electronics Letters, **34(14)**, pp. 1379–1380 (1998).

[67] A. Hyvärinen: "Independent component analysis in the presence of gaussian noise by maximizing joint likelihood", Neurocomputing, **22**, pp. 49–67 (1998).

[68] P. Hojen-Sorensen, O. Winther and L. K. Hansen: "Mean-field approaches to independent component analysis", Neural Computation, **14(4)**, pp. 889 – 918 (2002).

[69] K. B. Lee, T.-W. and R. Orglmeister: "Blind separation of nonlinear mixing models", IEEE International Workshop on Neural Networks for Signal Processing, pp. 406–415 (1997).

[70] R. Plomp: "Pitch of complex tones", J. Acoust. Soc. Am., **41**, pp. 1526–1533 (1967).

[71] R. J. Ristma: "Frequencies dominant in the perception of the pitch of complex sounds", J. Acoust. Soc. Am., **43**, pp. 191–198 (1967).

[72] R. M. Gray and D. L. Neuhoff: "Quantization", IEEE Trans. Inform. Theory, **44**, 6, pp. 2325–2383 (1998).

[73] A. Gersho and R. M. Gray: "Vector Quantization and Signal Compression", Kluwer Academic Publishers (1992).

[74] W. Y. Chan and A. Gersho: "Generalized product code vector quantization: A family of efficient techniques for signal compression", Digital Signal Processing, **4(2)**, pp. 95–126 (1994).

[75] W. R. Bennet: "Spectra of quantized signals", Bell Syst. Tech. J., **27**, pp. 446–472 (1948).

[76] P. F. Panter and W. Dite: "Quantizing distortion in pulse-count modulation with nonuniform spacing of levels", Proc. IRE, Vol. 39, pp. 44–48.

[77] P. W. Moo and D. L. Neuhoff: "Optimal compressor functions for multidimensional companding", IEEE International Symposium on Information Theory, Ulm (1997).

[78] V. K. Goyal: "Theoretical foundations of transform coding", IEEE Signal Processing Mag., **18(5)**, pp. 9–21 (2001).

[79] J. A. Bucklew: "A note on optimal multidimensional companders", IEEE Trans. Inform. Theory, **29(2)**, p. 279 (1983).

[80] A. Gersho: "Asymptotically optimal block quantization", IEEE Trans. Inform. Theory, **25**, 4, pp. 373–380 (1979).

[81] A. G. Tescher: "Transform image coding", Advances in electronics and electron physics, suppl.12, Academic Press, pp. 113–155 (1979).

[82] A. Buzo, A. H. Gray, R. M. Gray, Jr. and J. D. Markel: "Speech coding based upon vector quantization", IEEE Trans. Acoust. Speech Signal Process., **ASSP-28(5)**, pp. 562–574 (1980).

[83] H. P. Kramer and M. V. Mathews: "A linear coding for transmitting a set of correlated signals", IRE Trans. Inform. Theory, **IT-23**, pp. 41–46 (1956).

[84] J.-H. Huang and P. M. Schultheiss: "Block quantization of correlated gaussian random variables", IEEE Trans. Comm., **CS-11**, pp. 289–296 (1963).

付録

A.1 式 (4.31) の導出

期待対数尤度 $Q(A, c|\bar{A}, \bar{c})$ は次式で与えられる。

$$\begin{aligned} Q(A, c|\bar{A}, \bar{c}) &= -\frac{1}{T}\sum_{t=1}^{T}\left(\beta \int p(\mathbf{s}_t|\mathbf{x}_t, \bar{A}, \bar{c}) \left(g(\mathbf{x}_t; c) - A\mathbf{s}_t\right)^T \Sigma_t \left(g(\mathbf{x}_t; c) - A\mathbf{s}_t\right) d\mathbf{s}_t + \log Z(A, c)\right) \\ &= -\beta\frac{1}{T}\sum_{t=1}^{T}\left(g(\mathbf{x}_t; c)^T\Sigma_t g(\mathbf{x}_t; c) - 2g(\mathbf{x}_t; c)^T\Sigma_t A\left\langle \mathbf{s}_t\right\rangle + \mathrm{Tr}\left(\left\langle \mathbf{s}_t\mathbf{s}_t^T\right\rangle A^T\Sigma_t A\right)\right) + \log Z(A, c) \end{aligned}$$

ノイズが十分小さいとみなせるか、非線形関数 $g(\mathbf{x}_t; \theta)$ が線形な場合、β の掛かっていない項は無視できる。このとき、

$$\begin{aligned} F(\epsilon) &\equiv Q(A, c|\bar{A}, \bar{c}) - Q(\bar{A}, c|\bar{A}, \bar{c}) \\ &\approx -\beta\frac{1}{T}\sum_{t=1}^{T}\left(-2g(\mathbf{x}_t; c)^T\Sigma_t A\left\langle \mathbf{s}_t\right\rangle + \mathrm{Tr}\left(\left\langle \mathbf{s}_t\mathbf{s}_t^T\right\rangle A^T\Sigma_t A\right)\right) \\ &\quad +\beta\frac{1}{T}\sum_{t=1}^{T}\left(-2g(\mathbf{x}_t; c)^T\Sigma_t \bar{A}\left\langle \mathbf{s}_t\right\rangle + \mathrm{Tr}\left(\left\langle \mathbf{s}_t\mathbf{s}_t^T\right\rangle \bar{A}^T\Sigma_t \bar{A}\right)\right) \end{aligned}$$

となる。パラメータ A に $\bar{A} + \epsilon\Delta A$ を代入すると、

$$\begin{aligned} &= 2\beta\frac{1}{T}\sum_{t=1}^{T}\left(g(\mathbf{x}_t; c)^T\Sigma_t(\bar{A} + \epsilon\Delta A)\left\langle \mathbf{s}_t\right\rangle - g(\mathbf{x}_t; c)^T\Sigma_t\bar{A}\left\langle \mathbf{s}_t\right\rangle\right) \\ &\quad +\beta\frac{1}{T}\sum_{t=1}^{T}\left(-\mathrm{Tr}\left(\left\langle \mathbf{s}_t\mathbf{s}_t^T\right\rangle(\bar{A} + \epsilon\Delta A)^T\Sigma_t(\bar{A} + \epsilon\Delta A)\right) + \mathrm{Tr}\left(\left\langle \mathbf{s}_t\mathbf{s}_t^T\right\rangle \bar{A}^T\Sigma_t\bar{A}\right)\right) \\ &= -\epsilon^2\frac{\beta}{T}\sum_{t=1}^{T}\left(\mathrm{Tr}\left(\left\langle \mathbf{s}_t\mathbf{s}_t^T\right\rangle \Delta A^T\Sigma_t\Delta A\right)\right) + \epsilon\frac{2\beta}{T}\sum_{t=1}^{T}\left(g(\mathbf{x}_t; c)^T\Sigma_t\Delta A\left\langle \mathbf{s}_t\right\rangle - \mathrm{Tr}\left(\left\langle \mathbf{s}_t\mathbf{s}_t^T\right\rangle \Delta A^T\Sigma_t\bar{A}\right)\right). \end{aligned}$$

となり、式 (4.31) が導出できた。

A.2 係数 a と b の符号

勾配法は常に $F'(\epsilon = 0) > 0$ を満たすように勾配方向を定める。これは、係数 b が常に正であることを意味する。係数 a が正であるかどうかを示すのは、より複雑である。

$\mathrm{Tr}\left(\left\langle \mathbf{s}_t \mathbf{s}_t^T \right\rangle \Delta A^T \Sigma_t \Delta A\right)$ が正であれば係数 a は正である。そこで、$\mathrm{Tr}\left(\left\langle \mathbf{s}_t \mathbf{s}_t^T \right\rangle \Delta A^T \Sigma_t \Delta A\right)$ の正負を調べる。$\mathrm{Tr}\left(\left\langle \mathbf{s}_t \mathbf{s}_t^T \right\rangle \Delta A^T \Sigma_t \Delta A\right)$ は、2 行列 $\left\langle \mathbf{s}_t \mathbf{s}_t^T \right\rangle$ と $\Delta A^T \Sigma_t \Delta A$ の Hadamard 積をとった行列の要素をすべて足し合わせたものである。2 行列 P と Q の Hadamard 積を $P \odot Q$ で表す。p_{ij} と q_{ij} をそれぞれ行列 P と Q の (i, j) 成分とするとき、$P \odot Q$ の (i, j) 成分は $p_{ij} q_{ij}$ で表せる。以上の準備のもと以下の補題が成り立つことを示す。

補題 1

行列 U が (半) 正定値であるとき、その行列 U の全ての要素の和は非負である。

補題 2

2 つの $n \times n$ 行列 P と Q が (半) 正定値行列であるとき、それらの Hadamard 積 $U = P \odot Q$ もまた (半) 正定値行列である。

補題 3

行列 $\left\langle \mathbf{s}_t \mathbf{s}_t^T \right\rangle$ と $\Delta A^T \Sigma_t \Delta A$ はどちらも (半) 正定値行列である。

行列 U が (半) 正定値であるとき、その行列 U の全ての要素の和、$\mathbf{1}^T U \mathbf{1}$ は非負である。これは、補題 1 が成り立つことを示している。

次に、補題 2 を示す。2 つの $n \times n$ 行列 P と Q が (半) 正定値行列であるとき、行列 P と Q はそれぞれの下三角行列 L_p と L_q を用いて $P = L_p^T L_p$、 $Q = L_q^T L_q$ と表せる。a_{ij} と b_{ij} をそれぞれ下三角行列 L_p と L_q の (i, j) 成分であるとすると、$U = P \odot Q$ の (i, j) 成分 U_{ij} は以下のように計算される。

$$
\begin{aligned}
U_{ij} &= \left[(L_p^T L_p) \odot (L_q^T L_q) \right]_{ij} \\
&= \sum_{k=1}^{n} \sum_{l=1}^{n} a_{ki} a_{kj} b_{li} b_{lj}
\end{aligned}
$$

$$= \sum_{m=1}^{n^2} (\alpha_{m,i})\,(\alpha_{m,j}),$$

ここで、 $\alpha_{m,i} = a_{q(m)i} b_{r(m)i}$ であり、$q(m) = \lfloor m/n \rfloor$ 、$r(m) = m - q(m) \cdot n$ である。

$n^2 \times n$ 行列 M を (i,j) 成分が $\alpha_{i,j}$ である行列とすると、上式の関係を行列形式で書くことができる。

$$U = M^T M.$$

これは、行列 U が (半) 正定値であることを示している。

次に、補題 3 を示す。$\left\langle \mathbf{s}_t \mathbf{s}_t^T \right\rangle$ が (半) 正定値であることは明らか。$\Delta A^T \Sigma_t \Delta A$ は、Σ_t が正定値行列であることと、どの非ゼロベクトル $\mathbf{s}$ に対して $\mathbf{s}^T \Delta A^T \Sigma_t \Delta A \mathbf{s} = (\Delta A \mathbf{s})^T \Sigma_t (\Delta A \mathbf{s}) > 0$ が成り立つことから正定値行列であることがいえる。

補題 1、補題 2、補題 3 より係数 $\mathrm{Tr}\left(\left\langle \mathbf{s}_t \mathbf{s}_t^T \right\rangle \Delta A^T \Sigma_t \Delta A\right)$ は非負である。

業績リスト

論文

- ”学習による product code の設計”
  前田新一, 石井信
  電子情報通信学会論文誌, (2004), J87-A(3), 382-390.

国際会議

- ” An auditory system for efficient coding of natural sounds”,
  Shin-ichi Maeda and Shin Ishii,
  International Joint Conference on Neural Networks, Honolulu, (2002). 23-28.

- ”Optimization of product code”,
  Shin-ichi Maeda and Shin Ishii,
  International Conference on Neural Networks and Applications, (2004).

その他

- ”Separation of signal and noise from in vivo optical recording in Guinea pigs using independent component analysis”
  Shinichi Maeda, Shinji Inagaki, Hideo Kawaguchi, Wen-Jie Song,
  Neuroscience Letters. 302, (2001). 137-140.

- ”脳における予測と推定の仕組み”,
  石井 信, 前田 新一,
  Computer Today (解説記事). 111, (2002).

- ”冗長性圧縮に基づく聴覚モデル” ,
  前田新一, 石井 信,
  第 5 回情報論的学習理論ワークショップ, 山梨, (2002).