

# 修士論文

リザー計算を用いた  
予測符号化モデル

指導教員 河野 崇 教授

東京大学大学院工学系研究科  
電気系工学専攻

学籍番号 37-176458

氏名 田村 浩人

2019年1月31日 提出

# 論文要旨

近年神経科学において、感覚入力に対する内的知覚が脳内で形成される原理を、予測符号化 (Predictive Coding) という枠組みによって説明しようという試みが盛んである。特に PCRC (Predictive Coding with Reservoir Computing) モデルは、学習によって広いクラスの時系列を知覚できるようになる点で重要であるが、未学習の時系列を全く知覚できない、一括学習法を用いている、学習後のモデルが予測を行う原理が不明、等の問題を抱える。

そこで本研究では、これらの課題を克服した修正 PCRC モデルを構築し、異なる時間スケールの教師信号を用いて訓練した3つのネットワークについて、予測性能を調べた。また、これらのネットワークの予測メカニズムを、非線形動力学の観点から解析した。

さらに本研究では、修正 PCRC モデルに外部モジュールから送られる定常的なコンテキスト信号を組み込み、文脈依存型 PCRC モデルを構築した。このモデルにおいて、感覚入力が低次元の場合と高次元の場合の2つのネットワークについて予測性能を調べ、コンテキストに応じて処理を切り替える仕組みを非線形動力学の観点から解析した。また、受け取ったコンテキスト信号と感覚入力のタイプの組み合わせが訓練時と異なる場合 (コンテキストミスマッチ) についても解析を行った。

# 目次

第 1 章	序論	1
1.1	内的知覚	1
1.2	予測符号化理論	1
1.3	階層型予測符号化理論	2
1.4	PCRC(Predictive Coding with Reservoir Computing) モデル	4
第 2 章	従来の PCRC モデル	5
2.1	先行研究：Wacongne モデル	5
2.2	ネットワーク構造, ダイナミクス	5
2.3	学習則：Teacher forcing	6
2.4	Ara らの階層型 PCRC モデル	7
2.5	従来の PCRC モデルの課題, 本研究の目的	9
第 3 章	修正 PCRC モデル	10
3.1	ネットワーク構造, ダイナミクス	10
3.2	学習則：FORCE アルゴリズム	12
3.3	タスク	12
3.4	ネットワーク A：不連続変化する教師信号	15
3.5	ネットワーク B：ゆっくりと連続変化する教師信号	24
3.6	ネットワーク C：急速に連続変化する教師信号	29
第 4 章	文脈依存型 PCRC モデル	32
4.1	コンテキストとは	32
4.2	ネットワーク構造, ダイナミクス	33
4.3	感覚入力が低次元の場合	34
4.4	感覚入力が高次元の場合	39
第 5 章	結論, 考察	46
5.1	修正 PCRC モデル	46
5.2	文脈依存型 PCRC モデル	47

5.3	今後の課題 . . . . .	47
	謝辞	49
	参考文献	50
	発表文献	52

# 第 1 章

## 序論

### 1.1 内的知覚

神経科学や認知科学において、生物にとって主観的に見えているもの・聞こえているもの等を内的知覚と呼ぶ。内的知覚は感覚入力そのものではなく、感覚入力を統合し、外界の情報を自身の理解しやすい形に再構成したものであることが知られている。例えば、視覚入力は両眼の網膜において 2 つの異なる 2 次元画像として受け取られるが、ヒトの内的知覚においては 1 つの奥行きを持った 3 次元画像として見えている。聴覚においても、両耳から入ってきた物理的に異なる 2 つの音声情報が、ヒトの内的知覚としては音源の方向や距離感を伴った 1 つの音として聞こえている。また、感覚入力に付加されたノイズを除去したり、欠けている情報を自動的に補完したりする機構も内的知覚には備わっている。雑音を遮断して注意を向けた音声のみを聞き取れるカクテルパーティー効果 [1] や、遠近法の原理を用いて 2 次元の絵から 3 次元情報を読み取ろうとした際に起こる錯視 [2] 等がその代表例である。

このように、感覚入力と内的知覚の違いは広く認められており、行動学・心理学的なレベルでの知見は集積してきた。しかし神経科学の立場では、内的知覚が脳内で生成される機構については、その大部分が未だ明らかになっていない。

### 1.2 予測符号化理論

近年、感覚入力に対する内的知覚が脳内で生成される原理を、予測符号化 (Predictive coding) という枠組みで説明しようとする試みが、理論・実験の両面からなされている。詳細はモデル間で異なるものの、予測符号化理論では共通して以下の点を仮定する。

- 外界に存在する外部生成モデルが感覚入力を形成する。
- 脳内にも外部生成モデルを模した内部生成モデルが存在し、感覚入力の予測値を内的知覚として出力する。
- 感覚入力と予測値との誤差 (= 予測誤差) を基に内部生成モデルが最適化される。

脳において予測符号化が行われているという主張の根拠として、ミスマッチ陰性電位 (MMN: Mismatch Negativity) が挙げられる。MMN は脳波における事象関連電位 (ERP: Event-related Potential) の一種であり、先行する刺激の系列から逸脱するような刺激を受けた時に特徴的な反応を指す。(図 1.1 参照)。MMN は当初、聴覚刺激においてのみ報告されていた [3] が、近年では視覚刺激においても報告が相次いでいる [4]。予測符号化の立場では、この現象は「先行する刺激の系列によって内部生成モデルが最適化され、次の刺激を同様のものとして予測していたところにイレギュラーな刺激が入り、予測誤差が増大したため MMN が現れた」と解釈できる。実際、MMN の発生機構を説明する予測符号化モデルも提案されている [5][6]。

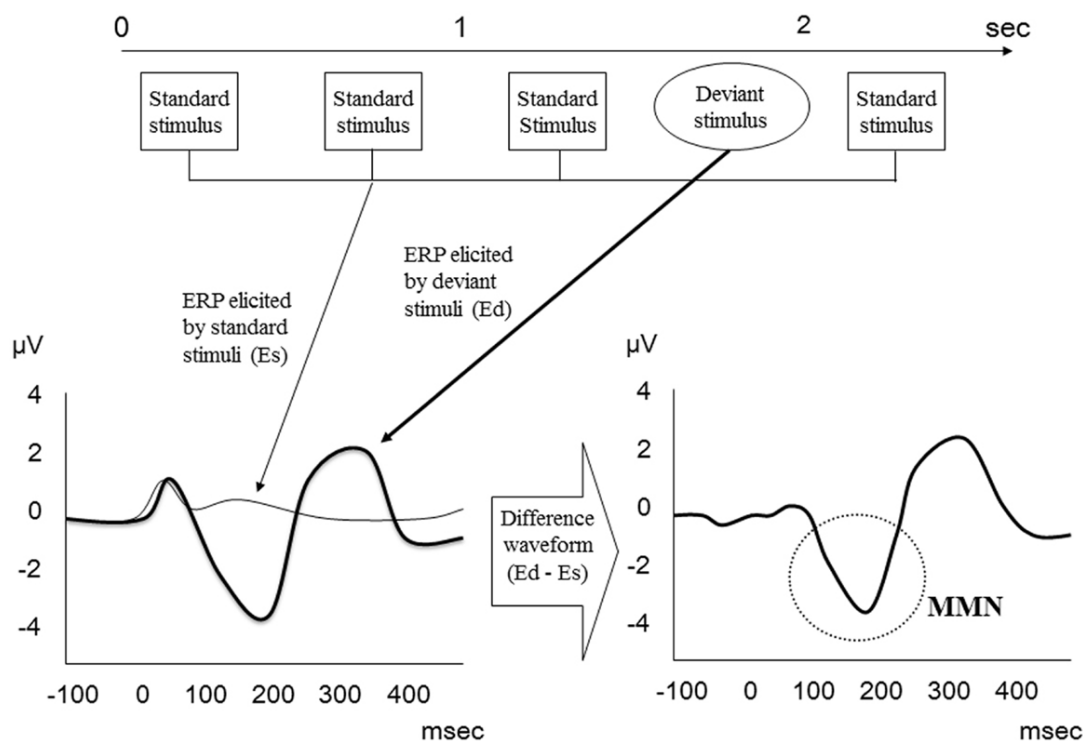


図 1.1 MMN の概要. [7] より引用.

### 1.3 階層型予測符号化理論

ヒトの視覚皮質や聴覚皮質において、低次領域から高次領域にかけての階層構造が存在することが知られている (例:  $V1 \rightarrow V2 \rightarrow V4 \rightarrow PIT \rightarrow CIT \rightarrow AIT$ ,  $V1 \rightarrow V2 \rightarrow MT/MIP \rightarrow MST/LIP$ , 一次聴覚野 $\rightarrow$ 二次聴覚野 $\rightarrow$ 三次聴覚野)。ここで、低次領域とはより感覚入力に近い領域を指す。これらの階層間の結合には順行性と逆行性の 2 種類があり、それぞれボトムアップ信号 (低次 $\rightarrow$ 高次) とトップダウン信号 (高次 $\rightarrow$ 低次) を伝達す

る。これら 2 方向の結合の非対称性が様々な面で報告されている。例えば,

- 解剖学的な非対称性：逆行性結合の方が数が多い [8].
- 機能的な非対称性：逆行性結合の方が神経修飾の影響を受けやすい [9].
- 動力的な非対称性：トップダウン信号の方が遅い時定数を持つ [10].

等である.

この階層間の双方向の信号の役割を説明するべく, 近年の予測符号化理論では特に階層型のモデルが多く提案されている [11][12][13]. 詳細はモデル間で異なるが, 階層型予測符号化理論では共通して以下の点を仮定する (図 1.2).

- トップダウン信号は, 低次領域の活動の予測値を表す.
- ボトムアップ信号は, トップダウン信号の伝えた予測と実際の低次領域の活動の間の予測誤差を表す.

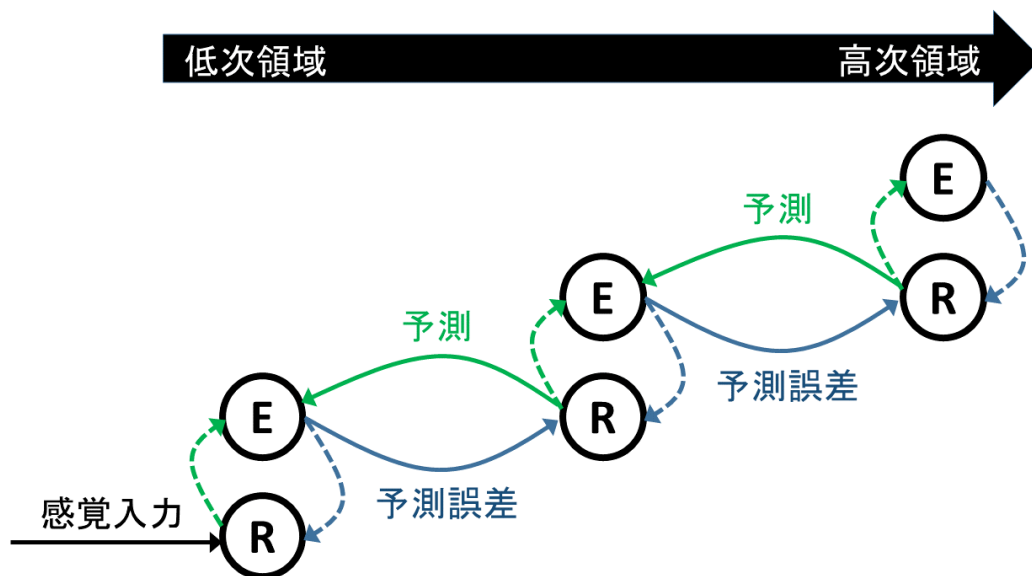


図 1.2 階層型予測符号化の概要.

階層型予測符号化モデルの先駆けとして知られるのが, Rao らによる視覚系のモデル [11] である. Rao らはまず視覚入力を出力する外部生成モデルに階層性を仮定し, 内部生成モデルにも同様の階層性を組み込むことで, 階層間における受容野の違いを再現し, 階層構造の役割を説明した. Rao モデルは階層型予測符号化の枠組みを確立した点で重要であるが, 静的な感覚入力を仮定しており, 動画像等の時系列入力には対応していない点が課題であった.

一方, Friston らによる自由エネルギー原理に基づく予測符号化モデル [12] は, 階層的な外部生成モデルから出力される時系列入力を対象としており, 階層的な内部生成モデルによって動的 EM アルゴリズム (DEM: Dynamical Expectation and Maximization) を実装している. Friston モデルの欠点として, 外部生成モデルのダイナミクスとその  $n$  階偏微分 ( $n = 1, 2, \dots$ ) が既知でなければ内部生成モデルを構成できない点が挙げられる.

なお, 近年機械学習やコンピュータービジョンなどの工学分野において注目を浴びている畳み込みニューラルネットワーク (CNN: Convolutional Neural Network)[14] も階層構造をもつ視覚野のモデルではあるが, CNN における階層間の結合は順行性結合のみの一方通行であり, このモデルによって脳の階層構造における強い逆行性結合の役割を説明することはできない.

## 1.4 PCRC(Predictive Coding with Reservoir Computing) モデル

Fukino らは PCRC(Predictive Coding with Reservoir Computing) モデル [15] を聴覚系のモデルとして提案した. PCRC モデルでは, リカレントニューラルネットワークの一種であるリザーバー計算 [16] によって内部生成モデルを実装しており, 動的な感覚入力に対する予測が可能であると同時に, 未知の外部生成モデルから出力された時系列に対しても学習によって予測が可能となる. すなわち, PCRC モデルは Rao モデルと Friston モデルのそれぞれの欠点を克服したモデルであるといえる.

また, その後 Ara らによって階層型の PCRC モデルも提案された [13]. このモデルは, Rao モデルに倣って階層構造を取り入れてみたものの, 1 層のみの PCRC モデルと比べて性能が落ちてしまっており, 試行段階にあるといえる.

Fukino らの PCRC モデル, および Ara らの階層型 PCRC モデルの実装や性能については次章で詳しく述べる.



## 第 2 章

# 従来の PCRC モデル

### 2.1 先行研究 : Wacongne モデル

Fukino らの PCRC モデル [15] は始め、聴覚系における MMN の発生機構を説明する Wacongne らの予測符号化モデル [6] を一般化したモデルとして提案された。Wacongne モデルは記憶痕跡層、予測層、予測誤差層、視床入力、の 4 つの部分から構成され、PCRC モデルでもこの基本構造を踏襲している。一方、Wacongne モデルはネットワークの素子としてスパイクニューロンを採用し、レセプターのモデルも生理学的な実態の再現を優先してしまっていたため、複雑な音型の処理が難しいという欠点があった。そこで、より抽象的なモデルである人工ニューラルネットワークモデルを用い、特に学習が容易なりザバー計算 [16] によるアプローチによって Wacongne モデルを簡素化し、複雑な音型にも対応できるようにしたのが Fukino らの PCRC モデルである。

### 2.2 ネットワーク構造, ダイナミクス

PCRC モデルでは、以下の差分方程式で表されるダイナミクスを用いる：

$$\begin{aligned} \mathbf{x}(n+1) = & \mathbf{x}(n) + \frac{\delta}{\tau} [-\mathbf{x}(n) \\ & + \phi(W^{REC}\mathbf{x}(n) + W^{FB}(\mathbf{z}(n) + \mathbf{r}(n)))], \end{aligned} \quad (2.1)$$

$$\mathbf{z}(n) = \phi(W^{OUT}\mathbf{x}(n)), \quad (2.2)$$

$$\mathbf{r}(n) = \max(0, \mathbf{d}(n) - \mathbf{z}(n)). \quad (2.3)$$

ここで、

$N \in \mathbb{N}$  : 記憶痕跡層の次元

$M \in \mathbb{N}$  : 予測層および視床入力の次元

$\tau \in \mathbb{R}$  : 時定数

$\delta \in \mathbb{R}$  : 微小な時間幅

$\mathbf{x}(n) := (x_1(n), \dots, x_N(n))^T \in \mathbb{R}^N$  : 時刻  $n \in \mathbb{Z}$  における記憶痕跡層の状態

$\mathbf{z}(n) \in \mathbb{R}^M$  : 時刻  $n$  における予測層の状態

$\mathbf{d}(n) \in \mathbb{R}^M$  : 時刻  $n$  における視床入力

$\mathbf{r}(n) \in \mathbb{R}^M$  : 時刻  $n$  における予測誤差

$\phi(\mathbf{x}) = (\phi(x_1), \dots, \phi(x_N))^T := (\tanh(x_1), \dots, \tanh(x_N))^T$  : ニューロンの非線形的な応答を表す活性化関数

とする. (図 2.1 参照)

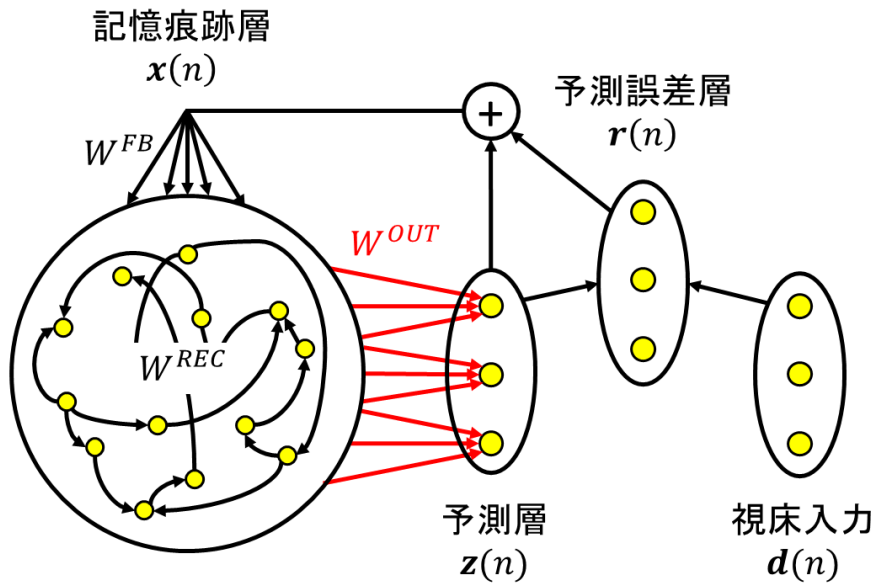


図 2.1 Fukino らの PCRC モデルのネットワーク構造の概念図. 学習中は赤い結合  $W^{OUT}$  のみが更新される.

$W^{REC} \in \mathbb{R}^{N \times N}$ ,  $W^{FB} \in \mathbb{R}^{N \times M}$ ,  $W^{OUT} \in \mathbb{R}^{M \times N}$  はそれぞれネットワークの結合係数行列である. PCRC モデルではリザーブ計算のアプローチ [16] に倣い,  $W^{REC}$ ,  $W^{FB}$  はランダムかつ疎な行列として固定し, 学習中は  $W^{OUT}$  のみを更新する.

## 2.3 学習則 : Teacher forcing

$W^{OUT}$  の学習には, Echo state network においてよく用いられる一括型学習法である Teacher forcing[16] を応用する. 具体的な学習法を以下に示す.

まず, ネットワークに学習させたい時間幅  $T_1 \in \mathbb{N}$  の視床入力  $\mathbf{d}(1), \dots, \mathbf{d}(T_1)$  を教師信号として用意する. 次に,  $W^{OUT}$  の全要素を 0 として初期化する ( $W^{OUT,(0)} = O$ ). 以下,  $m \in \mathbb{N}$  エポック目の  $W^{OUT}$  を  $W^{OUT,(m)}$  と表す. そして,  $W^{OUT,(m)}$  を用いて  $W^{OUT,(m+1)}$  を次の手順で計算する :

1.  $\mathbf{z}(n)$  の代わりに教師信号  $\mathbf{d}(n)$  を記憶痕跡層にフィードバックし, さらにノイズを加えたダイナミクスを  $n = 1, \dots, T_1$  の間進める (“Teacher forcing”). すなわち,

$$\begin{aligned} \mathbf{x}(n+1) = & \mathbf{x}(n) + \frac{\delta}{\tau} [-\mathbf{x}(n) \\ & + \phi(W^{REC}\mathbf{x}(n) + W^{FB}(\mathbf{d}(n) + \mathbf{r}(n)) + \sigma\xi(n))] \end{aligned} \quad (2.4)$$

で表されるダイナミクスを用いる. ここで  $\xi(n) \sim \mathcal{N}(0, 1)$  はネットワークに加えるノイズ,  $\sigma \in \mathbb{R}$  はその振幅を決める係数である.

2.  $0 < T_0 < T_1$  を満たす  $T_0$  をスレッシュホールド時刻として定め, 式 (2.4) のダイナミクスで得た  $\mathbf{x}(T_0), \dots, \mathbf{x}(T_1)$  を各行に持つ行列  $A \in \mathbb{R}^{(T_1-T_0+1) \times N}$  を用意する. また,  $\phi^{-1}(\mathbf{d}(T_0)), \dots, \phi^{-1}(\mathbf{d}(T_1))$  を各行に持つ行列  $B \in \mathbb{R}^{(T_1-T_0+1) \times M}$  を用意する.

3.  $\mathbf{d}(1), \dots, \mathbf{d}(T_1)$  に対してリッジ回帰を用いて  $W^{OUT,(m+1)}$  を求める. すなわち,

$$(W^{OUT,(m+1)})^T = (A^T A + \lambda I)^{-1} A^T B. \quad (2.5)$$

ここで,  $\lambda \in \mathbb{R}$  は正則化の係数,  $I \in \mathbb{R}^{N \times N}$  は単位行列, とする.

上の更新則を  $m = 0, 1, \dots$  として  $W^{OUT,(m)}$  が収束するまで繰り返すことにより, 目的の  $W^{OUT}$  を得る.

Fukino らは MMN を観測するための実験で用いられる音声入力を再現した  $M = 2$  次元の時系列を視床入力  $\mathbf{d}(n)$  としてとして用い, ネットワークに学習させた. 結果として, 予測誤差  $\mathbf{r} = \max(0, \mathbf{d} - \mathbf{z})$  の波形が MMN に対応することを確認した.

## 2.4 Ara らの階層型 PCRC モデル

Ara らはこの PCRC モデルを拡張し, 階層型 PCRC モデルを構築した. Ara モデルは低次領域と高次領域の 2 つの階層から成り, そのダイナミクスは以下の差分方程式で表される.

低次領域:

$$\begin{aligned} \mathbf{x}^l(n+1) = & \mathbf{x}^l(n) + \frac{\delta}{\tau} [-\mathbf{x}^l(n) \\ & + \phi(W^{l,REC}\mathbf{x}^l(n) + W^{l,FB}(\mathbf{z}^l(n) + \mathbf{z}^h(n)))] , \end{aligned} \quad (2.6)$$

$$\mathbf{z}^l(n) = \phi(W^{l,OUT}\mathbf{x}^l(n)), \quad (2.7)$$

$$\mathbf{r}^l(n) = \max(0, \mathbf{d}(n) - \mathbf{z}^l(n)). \quad (2.8)$$

高次領域 :

$$\mathbf{x}^h(n+1) = \mathbf{x}^h(n) + \frac{\delta}{\tau} \left[ -\mathbf{x}^h(n) + \phi \left( W^{h,REC} \mathbf{x}^h(n) + W^{h,FB} (\mathbf{z}^h(n) + \mathbf{r}^h(n)) \right) \right], \quad (2.9)$$

$$\mathbf{z}^h(n) = \phi(W^{h,OUT} \mathbf{x}^h(n)), \quad (2.10)$$

$$\mathbf{r}^h(n) = \max(0, \mathbf{r}^l(n) - \mathbf{z}^h(n)). \quad (2.11)$$

ここで、上付きの添え字  $l, h$  はそれぞれ低次領域, 高次領域を表し, その他の表記法は Fukino モデルのものに従うとする.

このネットワークの構造をまとめると図 2.2 のようになる.

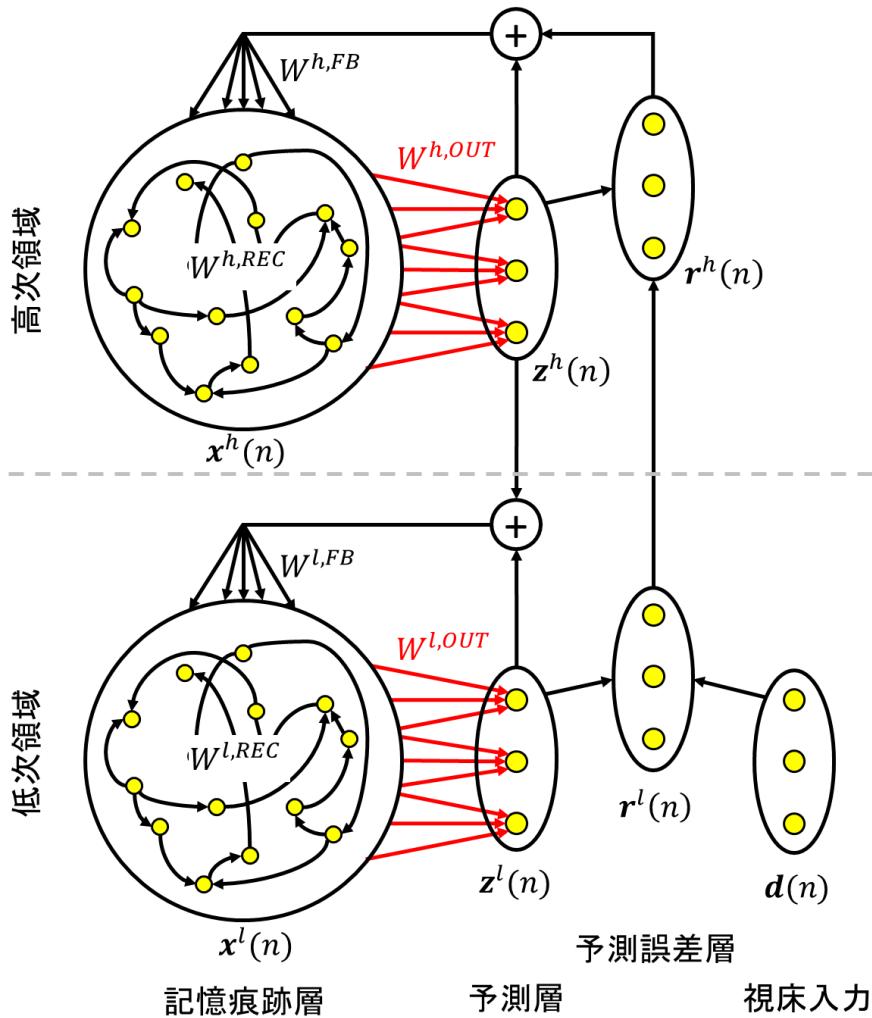


図 2.2 Ara らの階層型 PCRC モデルのネットワーク構造の概念図. 学習中は赤い結合  $W^{l,OUT}, W^{h,OUT}$  のみが更新される.

学習の際はまず低次領域のみから成る系を扱い、 $\mathbf{d}$  を教師信号として  $W^{l,OUT}$  に対して Teacher forcing による学習を行う。次に、低次領域と高次領域を接続した系を扱い、 $\mathbf{r}^l = \max(0, \mathbf{d} - \mathbf{z}^l)$  を教師信号として  $W^{h,OUT}$  に対して Teacher forcing による学習を行う。

Ara モデルにおいて階層構造を導入した目的は、長時間の相関を持つ複雑な音型の知覚が可能なモデルを構築するためであった。しかし、Ara らのシミュレーションによると、階層型 PCRC モデルは低次領域のみの PCRC モデルよりも予測が上手くいかなかったため、この階層化の試みは成功とはいえない。

## 2.5 従来の PCRC モデルの課題，本研究の目的

PCRC モデルは、Rao モデルや Friston モデルのそれぞれにおける欠点を克服した予測符号化モデルとして重要だが、以下に挙げた欠点を持つため、現状では脳における予測符号化を説明するモデルとしては不十分である。

Fukino モデルの欠点：

- 学習後のネットワークがブラックボックス化しており、予測のメカニズムについての数理的な説明がない。
- 未学習の時系列を全く知覚することができない。
- 学習則として採用している Teacher forcing は生理学的な実現性に乏しい。
- 予測値と予測誤差を  $\mathbf{z} + \mathbf{r} = \max(\mathbf{z}, \mathbf{d})$  という形で足してから記憶痕跡層にフィードバックしてしまっており、予測誤差を用いて予測を修正する予測符号化の枠組みからは外れてしまっている。

Ara モデルの欠点：

- 階層構造を導入する動機が曖昧。外部生成モデルの形を全く仮定していない。
- 階層構造を取り入れることにより性能が落ちてしまっている。
- 高次領域の出力  $\mathbf{z}^h$  を低次領域の予測誤差  $\mathbf{r}^l$  の予測値とすることの利点が不明瞭である。

そこで本研究では、PCRC モデルの正当性を高めるため、まず Fukino モデルの欠点を克服した修正 PCRC モデルを提案し、その性能と挙動の解析を行った (3 章)。さらに、外部モジュールから送られるコンテキスト信号を組み込んだ文脈依存型 PCRC モデルを提案し、そのコンテキストに応じた処理の切り替えについて解析を行った (4 章)。

## 第 3 章

# 修正 PCRC モデル

本章では、本研究の結果として、Fukino モデルの欠点を克服した修正 PCRC モデルの実装を説明する。また、訓練時の教師信号が、不連続変化する場合 (ネットワーク A)、ゆっくりと連続変化する場合 (ネットワーク B)、急速に連続変化する場合 (ネットワーク C)、の 3 つのネットワークを用意し、それぞれの予測性能を示す。さらに、各場合における学習後のネットワークの予測メカニズムについて、非線形動力学の観点から解析を行う。

### 3.1 ネットワーク構造, ダイナミクス

修正 PCRC モデルでは、以下の微分方程式で表されるダイナミクスを用いる：

$$\tau \dot{\mathbf{x}} = -\mathbf{x} + W^{REC} \mathbf{y} + W^{FB} \mathbf{z} + W^{IN} (\mathbf{d} - \mathbf{z}), \quad (3.1)$$

$$\mathbf{z} = W^{OUT} \mathbf{y}, \quad (3.2)$$

$$\mathbf{y} = \phi(\mathbf{x}). \quad (3.3)$$

ここで、

$N \in \mathbb{N}$  : リザバーの総ニューロン数

$M \in \mathbb{N}$  : 感覚入力および予測値の次元

$\tau \in \mathbb{R}$  : 時定数

$\mathbf{x}(t) := (x_1(t), \dots, x_N(t))^T \in \mathbb{R}^N$  : 時刻  $t \in \mathbb{R}$  におけるリザバーのニューロンの膜電位または活動度

$\mathbf{y}(t) \in \mathbb{R}^N$  : 時刻  $t$  におけるリザバーのニューロンの発火率

$\phi(\mathbf{x}) = (\phi(x_1), \dots, \phi(x_N))^T := (\tanh(x_1), \dots, \tanh(x_N))^T$  : ニューロンの非線形的な応答を表す活性化関数

$\mathbf{z}(t) \in \mathbb{R}^M$  : 時刻  $t$  におけるリザバーの出力または予測値

$\mathbf{d}(t) \in \mathbb{R}^M$  : 時刻  $t$  における感覚入力

とする。(図 3.1 参照).

$W^{REC} \in \mathbb{R}^{N \times N}$ ,  $W^{FB} \in \mathbb{R}^{N \times M}$ ,  $W^{IN} \in \mathbb{R}^{N \times M}$ ,  $W^{OUT} \in \mathbb{R}^{M \times N}$  はそれぞれシナプス強度に対応するネットワークの結合係数行列である。本モデルではリザーバ計算 [16] を用いるため、学習の間  $W^{REC}$ ,  $W^{FB}$ ,  $W^{IN}$  は固定され、 $W^{OUT}$  のみが更新される。 $W^{REC}$  の各要素はガウス分布  $\mathcal{N}(0, p^2/N)$  から i.i.d. にサンプリングされ、 $W^{FB}$ ,  $W^{IN}$  の各要素は一様分布  $\mathcal{U}[-1, 1]$  から i.i.d. にサンプリングされる。また、 $W^{OUT}$  の初期値は零行列とする。

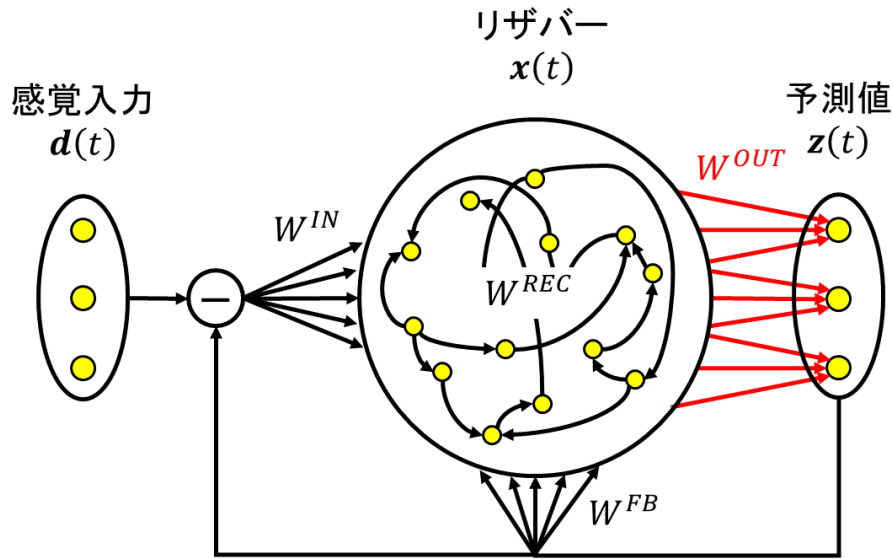


図 3.1 修正 PCRC モデルのネットワーク構造の概念図。学習中は赤い結合  $W^{OUT}$  のみが更新される。

本モデルの式 (3.1) における予測誤差項  $W^{IN}(\mathbf{d} - \mathbf{z})$  は、Fukino モデルの式 (2.1) における予測誤差項  $W^{FB} \max(0, \mathbf{d} - \mathbf{z})$  に比べてより自然な形である。また、Fukino モデルでは  $W^{FB}(\mathbf{z} + \max(0, \mathbf{d} - \mathbf{z})) = W^{FB} \max(\mathbf{d}, \mathbf{z})$  のように予測項と予測誤差項の和がリザーバにフィードバックされていたのに対し、本モデルでは  $W^{FB}\mathbf{z} + W^{IN}(\mathbf{d} - \mathbf{z})$  という形で予測項と予測誤差項は独立してフィードバックされている。従って、本モデルは厳密に「予測誤差入力によって駆動している」といえる。

式 (3.1) のダイナミクスを数値的にシミュレーションする際は、オイラー法によって得られる以下の離散時間版のダイナミクスを用いる：

$$\mathbf{x}(n+1) = \mathbf{x}(n) + \frac{\delta}{\tau} [-\mathbf{x}(n) + W^{REC}\mathbf{y}(n) + W^{FB}\mathbf{z}(n) + W^{IN}(\mathbf{d}(n) - \mathbf{z}(n))]. \quad (3.4)$$

ここで、 $n \in \mathbb{Z}$  は離散時間ステップ、 $\delta$  は微小時間間隔とし、他の表記は式 (3.1) に従う。数値シミュレーション時には、本論文を通して  $N = 1000$ ,  $p = 1.2$ ,  $\tau = 0.1[\text{sec}]$ ,  $\delta = 0.01[\text{sec}]$  を用いる。また、本章を通して、簡単のため  $M = 2$  とする。

## 3.2 学習則 : FORCE アルゴリズム

$W^{OUT}$  の学習には, 感覚入力  $\mathbf{d}$  を教師信号として, リザバー計算における強力な逐次型学習則である FORCE(Fast Order Reduced and Controlled Error) アルゴリズム [17] を用いる. 以下, 離散時間版のダイナミクス (式 (3.4)) を考え, 時刻  $n$  における  $W^{OUT}$  の値を  $W^{OUT}(n)$  とする. FORCE アルゴリズムは再帰的最小二乗フィルタに基づいており, 次の損失関数  $\text{Loss}(W^{OUT}(n))$  の最小化を考えることで導出される :

$$\text{Loss}(W^{OUT}(n)) := \sum_{j=0}^n \|\mathbf{e}(j)\|^2, \quad (3.5)$$

$$\text{where } \mathbf{e}(j) := W^{OUT}(j-1)\mathbf{y}(j) - \mathbf{d}(j). \quad (3.6)$$

結局, FORCE アルゴリズムの更新則は以下となる :

$$\mathbf{s}(n) := \frac{P(n-1)\mathbf{y}(n)}{1 + \mathbf{y}^T(n)P(n-1)\mathbf{y}(n)}, \quad (3.7)$$

$$P(n) = P(n-1) - \mathbf{s}(n)\mathbf{y}^T(n)P(n-1), \quad (3.8)$$

$$W^{OUT}(n) = W^{OUT}(n-1) - \mathbf{e}(n)\mathbf{s}^T(n). \quad (3.9)$$

ここで,  $P(n) \in \mathbb{R}^{N \times N}$  の初期値は

$$P(0) = \frac{1}{\alpha}I \quad (\alpha \in \mathbb{R}, I \in \mathbb{R}^{N \times N} \text{ は単位行列}) \quad (3.10)$$

で与えられる.

本論文を通して, 数値シミュレーションでは  $\alpha = 0.02$  を用いる.

## 3.3 タスク

### 3.3.1 訓練時

訓練時は, 時間スケールの異なる 3 つの教師信号を用意し, それぞれ異なるネットワークの学習に用いた. 訓練後のネットワークをそれぞれネットワーク A, B, C と呼ぶ.

- ネットワーク A : 不連続変化する教師信号

$N_D = 1000$  を試行回数とし,  $i$  番目の試行 ( $i = 1, \dots, N_D$ ) において定ベクトル  $\mathbf{d}^i$  を感覚入力および教師信号としてネットワークに与える. 各  $\mathbf{d}^i$  は  $0.2[\text{sec}]$  の間提示され, その成分は一様分布  $\mathcal{U}[1, 2]$  から i.i.d. にサンプリングされる. (図 3.2 参照.)



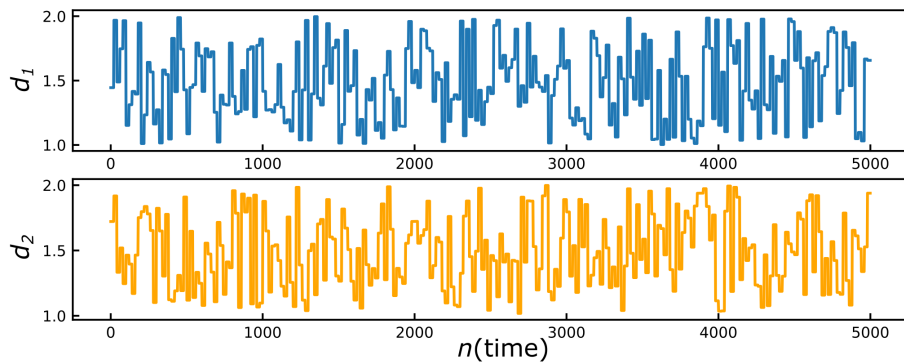


図 3.2 訓練時のネットワーク A に与えられる教師信号  $\mathbf{d}(t)$  の一部.

- ネットワーク B: ゆっくりと連続変化する教師信号

教師信号となる感覚入力  $\mathbf{d}(t) = (d_1(t), d_2(t))^T$  として, 区間  $[1, 2]^2$  の間に収まる, ネットワークの時定数に比べて遅い周期を持つサイン波を用意した (図 3.3 参照):

$$d_1(t) = 0.5 \cdot \sin\left(0.002 \cdot \frac{t}{\delta}\right) + 1.5, \quad (3.11)$$

$$d_2(t) = 0.5 \cdot \sin\left(0.003 \cdot \frac{t}{\delta}\right) + 1.5. \quad (3.12)$$

このとき,  $d_1(t)$  と  $d_2(t)$  で周期がずれるように設定しているため, これらの値の間に相関がなくなることに注意する. 訓練時は, 50[sec] の間学習を行った. ネットワークは予測誤差  $\mathbf{d}(t) - \mathbf{z}(t)$  から元の感覚入力  $\mathbf{d}(t)$  を逐次的に復号化することを要求される.

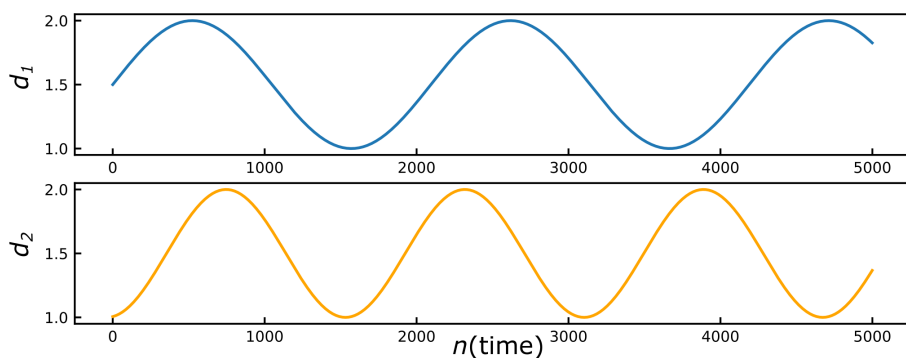


図 3.3 訓練時のネットワーク B に与えられる教師信号  $\mathbf{d}(t)$ .

- ネットワーク C : 急速に連続変化する教師信号

教師信号となる感覚入力  $\mathbf{d}(t) = (d_1(t), d_2(t))^T$  として, 区間  $[1, 2]^2$  の間に収まる, ネットワークの時定数と同程度の周期を持つサイン波を用意した (図 3.4 参照) :

$$d_1(t) = 0.5 \cdot \sin\left(0.02 \cdot \frac{t}{\delta}\right) + 1.5, \quad (3.13)$$

$$d_2(t) = 0.5 \cdot \sin\left(0.03 \cdot \frac{t}{\delta}\right) + 1.5. \quad (3.14)$$

このとき,  $d_1(t)$  と  $d_2(t)$  の間に相関がなくなること注意到する. 訓練時は, 50[sec] の間学習を行った.

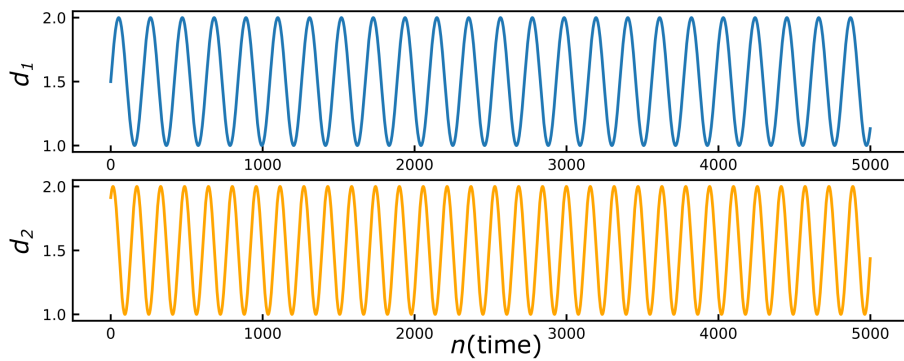


図 3.4 訓練時のネットワーク C に与えられる教師信号  $\mathbf{d}(t)$ .

### 3.3.2 テスト時

テスト時は, 訓練後のネットワーク A, B, C に対して,  $W^{OUT}$  を固定し, 時間スケールの異なる 3 つの感覚入力それぞれに対する応答を調べた. テスト時の感覚入力をそれぞれ, テスト入力 A, B, C と呼ぶ.

- テスト入力 A : 不連続変化する感覚入力

テスト入力 A の基本的な設定は訓練時にネットワーク A に与えた教師信号に準じた. ただし, 値の収束の有無を確認するため, 各試行の長さは 5.0[sec] と長めに設定した. テスト入力 A はランダムに生成されるため, ネットワーク A に対してもテスト入力 A は未学習の時系列となることに注意されたい.

- テスト入力 B : ゆっくりと連続変化する感覚入力テスト入力 B として, ネットワーク B の訓練時に用いた教師信号と同程度の周期のサイン波を用意した :

$$d_1(t) = 0.5 \cdot \sin\left(0.0021 \cdot \frac{t}{\delta}\right) + 1.5, \quad (3.15)$$

$$d_2(t) = 0.5 \cdot \sin\left(0.002 \cdot \frac{t}{\delta}\right) + 1.5. \quad (3.16)$$

この際, ネットワーク B に対してもテスト入力 B は未学習の時系列となることに注意されたい.

- テスト入力 C : 急速に連続変化する感覚入力  
テスト入力 C として, ネットワーク C の訓練時に用いた教師信号の  $d_1(t), d_2(t)$  の値を入れ替えたものを用意した. すなわち,

$$d_1(t) = 0.5 \cdot \sin\left(0.03 \cdot \frac{t}{\delta}\right) + 1.5, \quad (3.17)$$

$$d_2(t) = 0.5 \cdot \sin\left(0.02 \cdot \frac{t}{\delta}\right) + 1.5. \quad (3.18)$$

この際, ネットワーク C に対してもテスト入力 C は未学習の時系列となることに注意されたい.

以下では, ネットワーク A, B, C による予測の性能とそのメカニズムについて述べる.

## 3.4 ネットワーク A : 不連続変化する教師信号

### 3.4.1 性能

テスト入力 A に対するネットワーク A の応答を図 3.5 に示す. この図を見ると, ネットワーク A によるテスト入力 A の予測はほとんど完全に成功している. 図の  $\mathbf{d} - \mathbf{z}$  の行に注目すると, 各試行  $i$  において, 予測誤差  $\mathbf{d}^i - \mathbf{z}(t)$  は始め鋭いパルス状の信号としてリザーブに入力されるが, その後直ちに減衰し, ネットワークは固定点に収束しているように見える. 試行  $i$  の最後における  $\mathbf{x}$  を以下では  $\bar{\mathbf{x}}^i$  と呼ぶ.

次に, テスト入力 B に対するネットワーク A の応答を図 3.6 に示す. この図を見ると, ネットワーク A はテスト入力 B を 0.1 以内の予測誤差で予測できていることが分かる.

最後に, テスト入力 C に対するネットワーク A の応答を図 3.7 に示す. この図を見ると, ネットワーク A はテスト入力 C を 0.1 以内の予測誤差で予測できていることが分かる. また, 予測誤差  $\mathbf{d} - \mathbf{z}$  の行を見ると, 0 付近で連続的に変化している様子が確認できる.

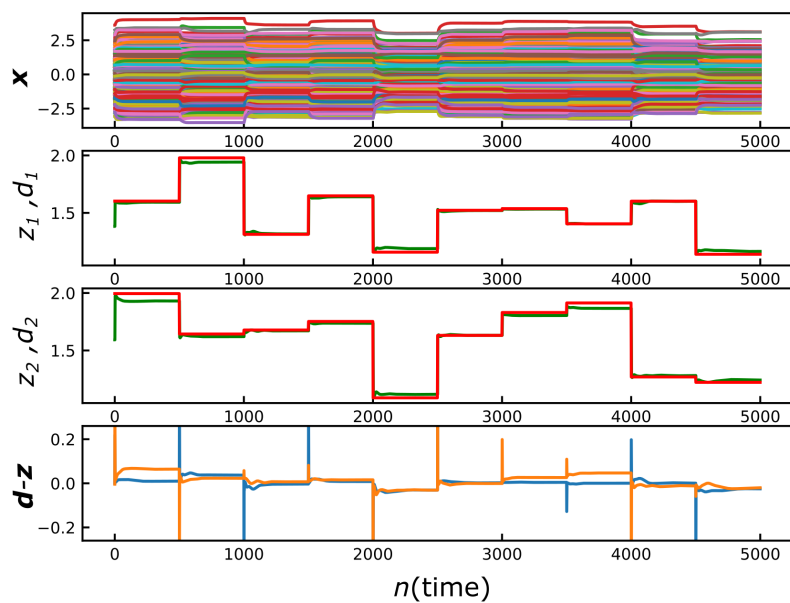


図 3.5 ネットワーク A のテスト入力 A に対する応答. 2,3 行目の赤いプロットは感覚入力  $d(t)$ , 2,3 行目の緑のプロットは予測  $z(t)$  を表す.

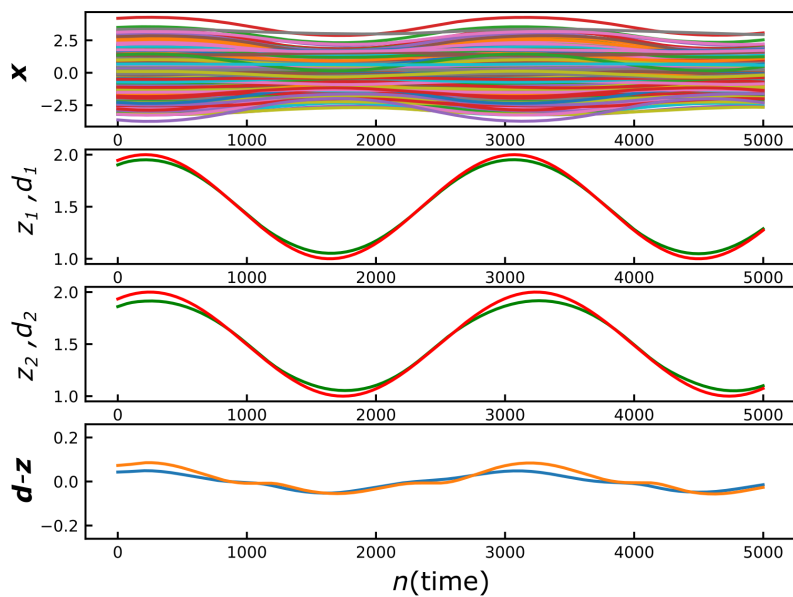


図 3.6 ネットワーク A のテスト入力 B に対する応答. 2,3 行目の赤いプロットは感覚入力  $d(t)$ , 2,3 行目の緑のプロットは予測  $z(t)$  を表す.

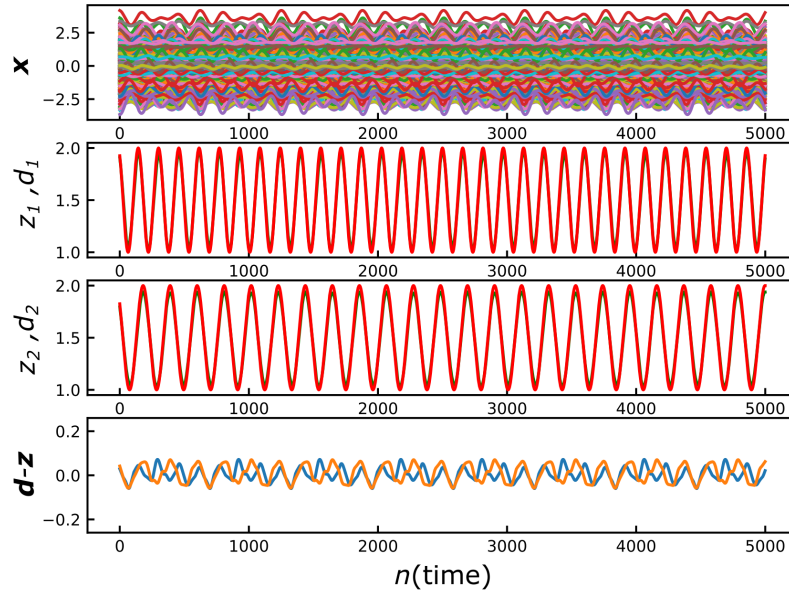


図 3.7 ネットワーク A のテスト入力 C に対する応答. 2,3 行目の赤いプロットは感覚入力  $d(t)$ , 2,3 行目の緑のプロットは予測  $z(t)$  を表す.

### 3.4.2 非線形ダイナミクス解析

前節で示したネットワーク A の挙動の背後にあるメカニズムを明らかにするため、本節では非線形動力学の観点から解析を行う。

まず、テスト入力 A の各試行  $i$  の最後における固定点  $\bar{x}^i (i = 1, 2, \dots)$  の位相空間における配置を、主成分分析 (PCA: Principal Component Analysis) を用いて視覚化すると、図 3.8 の青い点群のようになった。この図より、これらの固定点  $\bar{x}^i$  は位相空間において 1 つの 2 次元多様体上に分布していることが分かる。従って、この 2 次元多様体上での議論をする際は、2 次元の PCA 空間上での視覚化で十分である。また、図 3.8 における 1 試行あたりの軌道 (黄色いプロット) を見ると、試行の始めに軌道は 2 次元多様体上から急激に引き離されるが、その後また 2 次元多様体上に吸引されていく様子が確認できる。

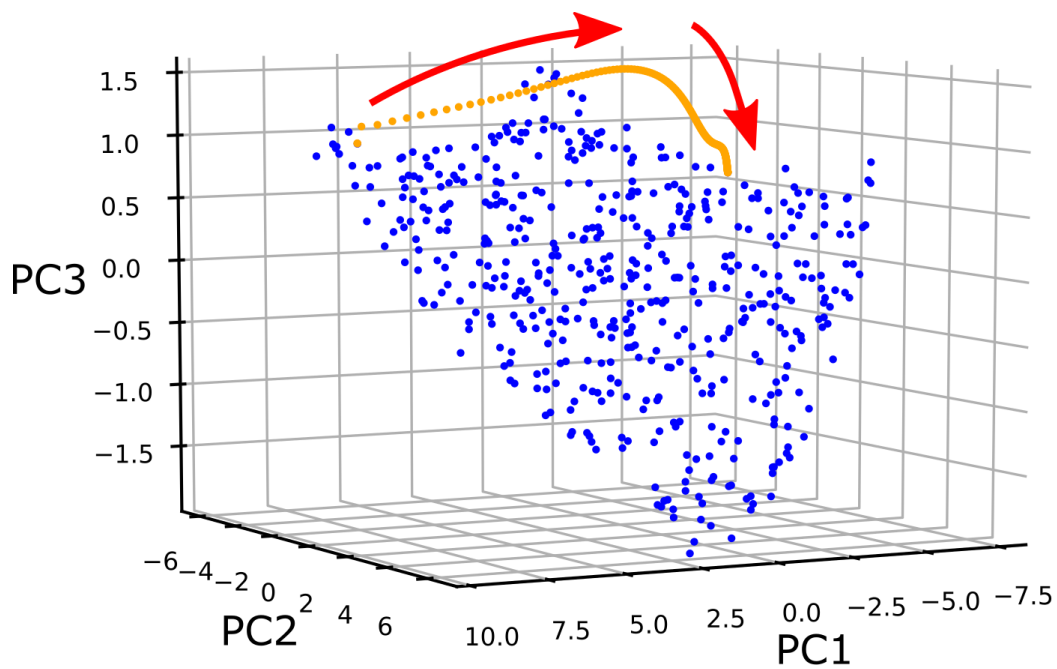


図 3.8 ネットワーク A の位相空間における固定点の位置 (青) と 1 試行あたりの軌道 (黄色).

次に、2次元多様体上の固定点  $\bar{x}$  の配置とその点における予測値  $\bar{z} := W^{OUT} \phi(\bar{x})$  の値の関係を視覚化すると、図 3.9, 図 3.10 のようになった。これらの図から、2次元多様体上において予測値  $\bar{z}$  の値は連続的に配置されていることが分かる。さらに、図 3.8 における 1 試行あたりの軌道 (黄色いプロット) と合わせて考えると、各試行  $i$  の最初と最後を比較したときの軌道の変位  $\bar{x}^i - \bar{x}^{i-1}$  が、予測値の変化量  $\mathbf{d}^i - \mathbf{d}^{i-1}$  に対応していることが確認できる。

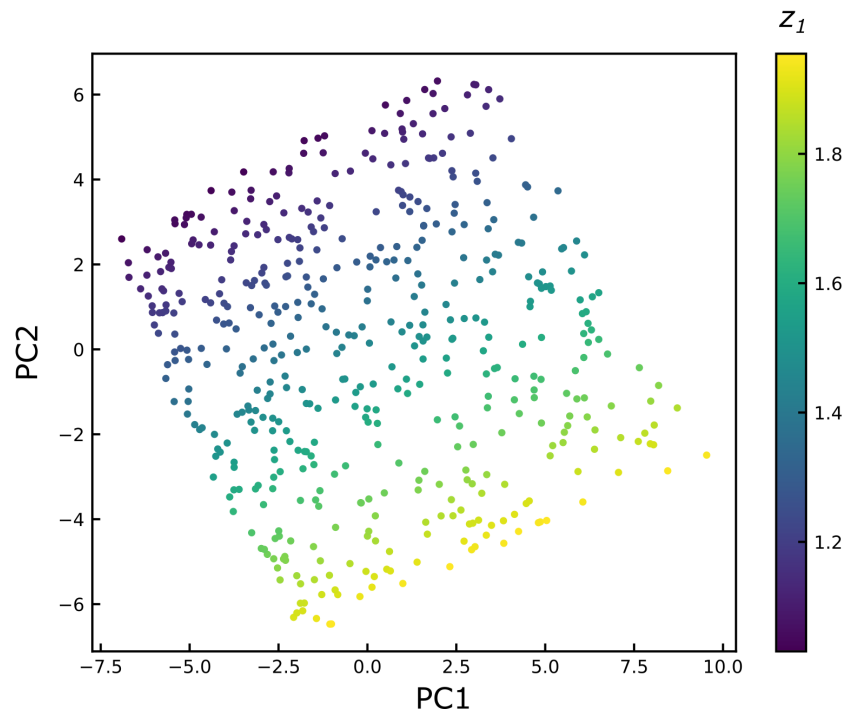


図 3.9 ネットワーク A の 2 次元多様体上の固定点  $\bar{x}$  の配置と予測値  $\bar{z}_1$  の関係.

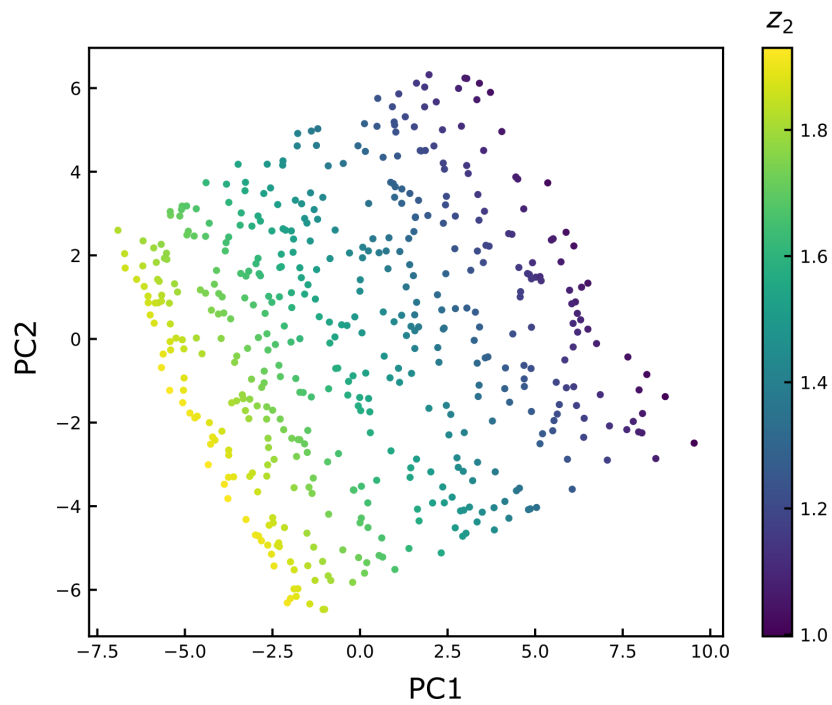


図 3.10 ネットワーク A の 2 次元多様体上の固定点  $\bar{x}$  の配置と予測値  $\bar{z}_2$  の関係.

以下では、この固定点で構成される2次元多様体について、より詳細な性質を調べる。

式(3.1)における予測誤差項  $W^{IN}(\mathbf{d} - \mathbf{z})$  は各試行の始め以外はほぼ0とみなすことができるため、以下ではこれを外力として扱い、ネットワーク自身のダイナミクスから分離して考える。すなわち、次の方程式で表されるダイナミクスについて解析を行う：

$$\tau \dot{\mathbf{x}} = -\mathbf{x} + W^{REC} \mathbf{y} + W^{FB} \mathbf{z} \quad (3.19)$$

式(3.1)の力学系はネットワークの内部状態  $\mathbf{x}(t), \mathbf{z}(t)$  とは関係なく時間変化する感覚入力  $\mathbf{d}(t)$  が組み込まれており、直接解析するのは困難であったが、式(3.19)は感覚入力  $\mathbf{d}(t)$  とは独立した系であるため、直接解析することが比較的容易である。

まず、2次元多様体の点  $\bar{\mathbf{x}}^i (i = 1, 2, \dots)$  が本当に数理的に固定点とみなせるかを調べるため、Sussillo と Barak による解析手法 [18] に倣い、次の疑似エネルギー関数  $q(\mathbf{x})$  を用いる：

$$q(\mathbf{x}) := \frac{1}{2} \|\dot{\mathbf{x}}\|^2 = \frac{1}{2\tau^2} \|\mathbf{x} - W^{REC} \mathbf{y} - W^{FB} \mathbf{z}\|^2. \quad (3.20)$$

ここで、「 $q(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x}$  は固定点」であり、 $q(\mathbf{x}) \simeq 0$  であれば  $\mathbf{x}$  は実質的に固定点とみなせることに注意する。

2次元多様体上の点  $\bar{\mathbf{x}}^i (i = 1, 2, \dots)$  における疑似エネルギー  $q(\bar{\mathbf{x}}^i)$  の値を視覚化すると、図3.11のようになった。この図より、2次元多様体上の中心部は  $q(\mathbf{x})$  の値が小さく、実質的に固定点とみなせるが、端の部分は  $q(\mathbf{x})$  の値が大きく、正確には固定点とは呼べないことが分かる。後者の場合でもテスト入力 A に対して固定点のように振舞っているのは、微小な予測誤差  $\mathbf{d} - \mathbf{z}$  によってつり合いがとれているためと考えられる。



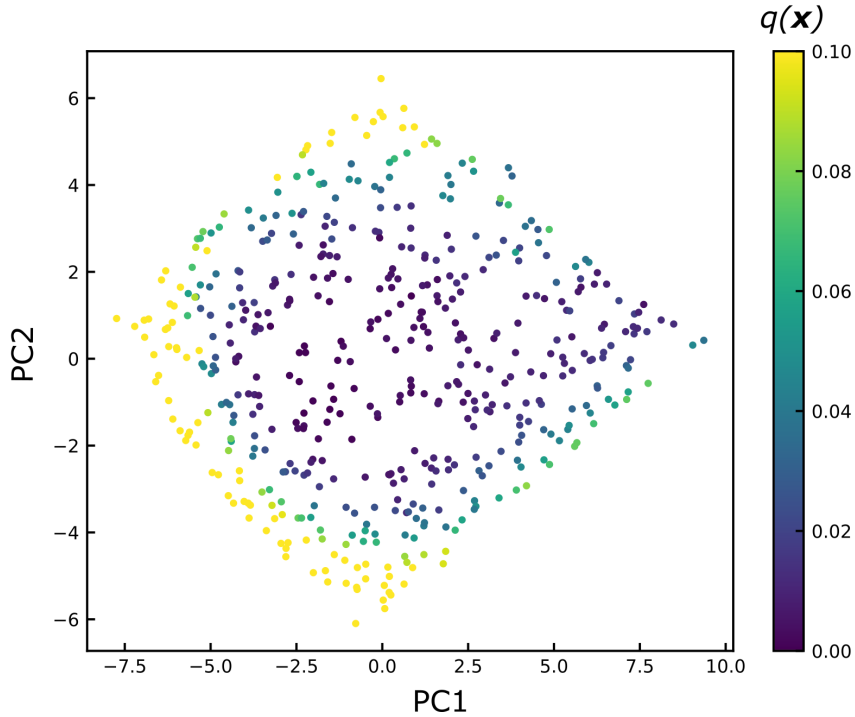


図 3.11 ネットワーク A の 2 次元多様体上の固定点  $\bar{\mathbf{x}}$  の位置と疑似エネルギー  $q(\bar{\mathbf{x}})$  の関係. カラースケールは  $q(\mathbf{x})$  の上限を 0.1 として表示している.

最後に、この無数の固定点から成る 2 次元多様体の安定性を調べるため、固定点  $\bar{\mathbf{x}}$  の近傍において式 (3.19) を線形近似し、摂動  $\delta\mathbf{x} := \mathbf{x} - \bar{\mathbf{x}}$  に関する次のダイナミクスを得た：

$$\dot{\delta\mathbf{x}} = \frac{1}{\tau} [-I + (W^{REC} + W^{FB}W^{OUT})R'(\bar{\mathbf{x}})] \delta\mathbf{x}, \quad (3.21)$$

$$\text{where } R'_{ij} := \delta_{ij}\phi'(\bar{x}_i). \quad (3.22)$$

2 次元多様体上の固定点  $\bar{\mathbf{x}}^i (i = 1, 2, \dots)$  におけるヤコビ行列の固有値をプロットすると、ほとんどの場合図 3.12 のように分布し、その実部は全て負であった。さらに、2 次元多様体上においてその固有値の実部のうち最大のを視覚化すると、図 3.13 のようになった。これらの図から、 $\bar{\mathbf{x}}^i (i = 1, 2, \dots)$  のほとんどが安定固定点であり、これらが 2 次元多様体上に分布することで 1 つのアトラクターのように振舞っていることが分かる。すなわち、位相空間において、この 2 次元多様体は近傍の全ての軌道を吸引する。

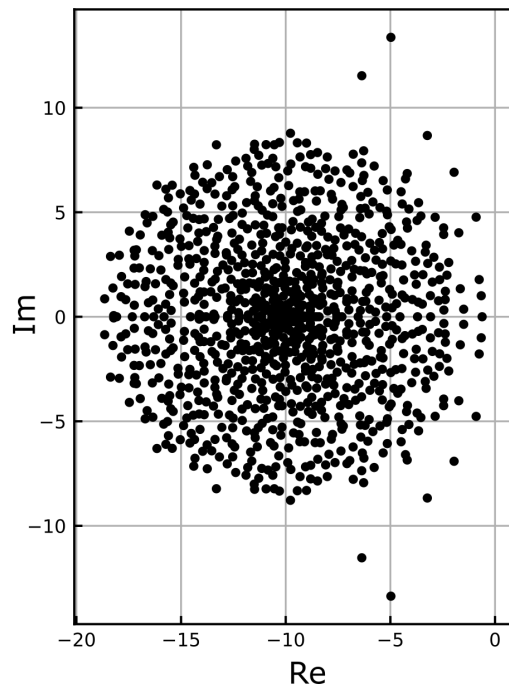


図 3.12 ネットワーク A の 2 次元多様体上の固定点のヤコビ行列における固有値スペクトラムの典型例.

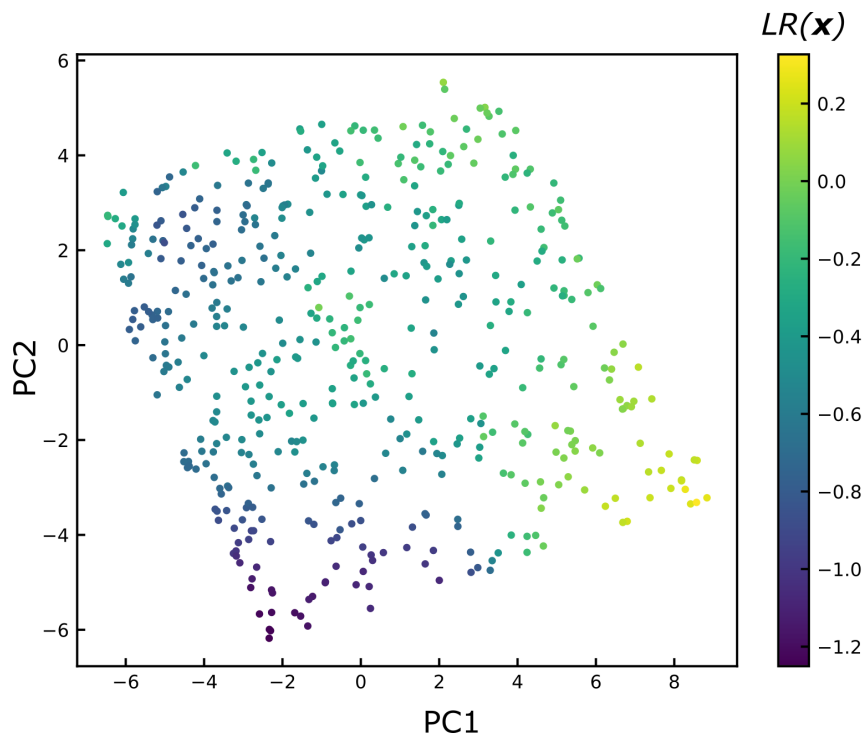


図 3.13 ネットワーク A の 2 次元多様体上の固定点  $\bar{x}$  の位置とヤコビ行列の固有値の最大実部  $LR(\bar{x})$  の関係.

しかし、これまで示してきた通り、2次元多様体を構成する点には完全な安定固定ではないものも含まれているため、この2次元多様体上にも極めて遅いフローが存在する。従って、図 3.8 に示されるテスト入力 A の各試行  $i$  における軌道の典型的な振舞いは、以下のように説明できる：

1. 試行  $i-1$  の最後の時点では  $z \simeq \mathbf{d}^{i-1}$  で、軌道は2次元多様体上にある。
2. 試行  $i$  の始めに、パルス状の予測誤差入力  $\mathbf{d}^i - z (\simeq \mathbf{d}^i - \mathbf{d}^{i-1})$  によって軌道は2次元多様体上から急激に引き離される。
3. 次に、軌道は2次元多様体に吸引される。
4. 最後に、2次元多様体上の遅いフローに流されて、軌道はゆっくりと特定の安定固定点に吸引される ( $z \simeq \mathbf{d}^i$ )。

次に、位相空間における2次元多様体とテスト入力 B に対する応答時の  $\mathbf{x}(t)$  の軌道の関係を視覚化すると、図 3.14 のようになった。この図から、 $\mathbf{x}(t)$  の軌道は多様体アトラクター上に完全に落ち着くことはなく、その周りを旋回している様子が確認できる。これらのことから、ネットワーク A において、テスト入力 B の予測は、多様体アトラクターによる吸引力と予測誤差  $\mathbf{d}(t) - z(t)$  による駆動力とのバランスにより実現されていることが分かる。テスト入力 C に対する予測にも同様のメカニズムが働いていると考えられる。

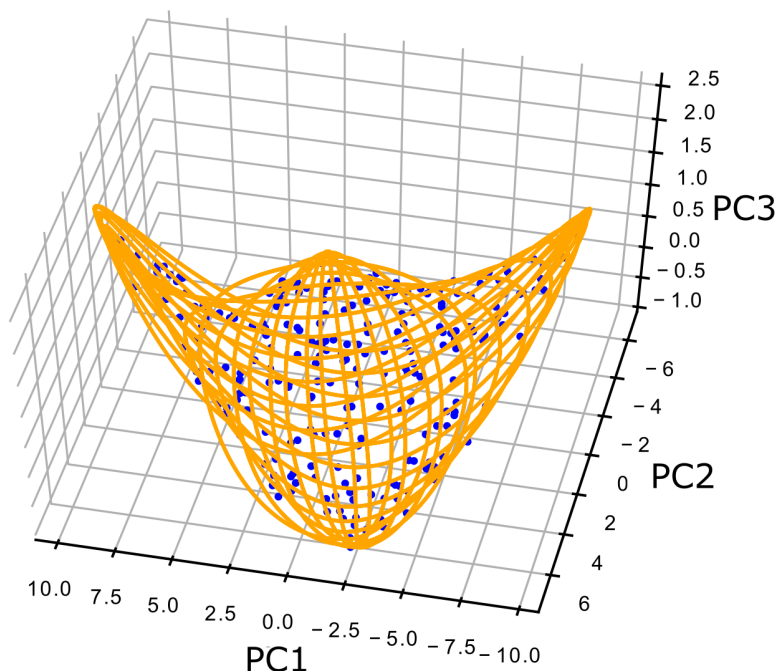


図 3.14 ネットワーク A の位相空間における2次元多様体 (青) とテスト入力 B に対する応答  $\mathbf{x}(t)$  の軌道 (黄色)。

## 3.5 ネットワーク B : ゆっくりと連続変化する教師信号

### 3.5.1 性能

テスト入力 A に対するネットワーク B の応答を図 3.15 に示す。この図が示す通り、ネットワーク B によるテスト入力 B の予測はほとんど完全に成功している。また、図 3.5 と比較すると、各試行の最後における予測誤差  $d - z$  のノルムはネットワーク A よりもネットワーク B の方が小さく、最終的な予測の正確さはより高いといえる。一方、どちらも各試行の最初にはパルス的な予測誤差を示すが、ネットワーク A よりもネットワーク B の方が固定点への収束が遅いことが確認できる。

次に、テスト入力 B に対するネットワーク B の応答を図 3.16 に示す。この図から、ネットワーク B はテスト入力 B を 0.01 以内の予測誤差で予測できていることが分かる。これはネットワーク A の時に比べて 1 割程度の誤差である。

最後に、テスト入力 C に対するネットワーク B の応答を図 3.17 に示す。この図から、ネットワーク B はテスト入力 C を 0.25 以内の予測誤差で予測できていることが分かる。すなわち、ネットワーク A に比べてテスト入力 C の予測精度は劣る。

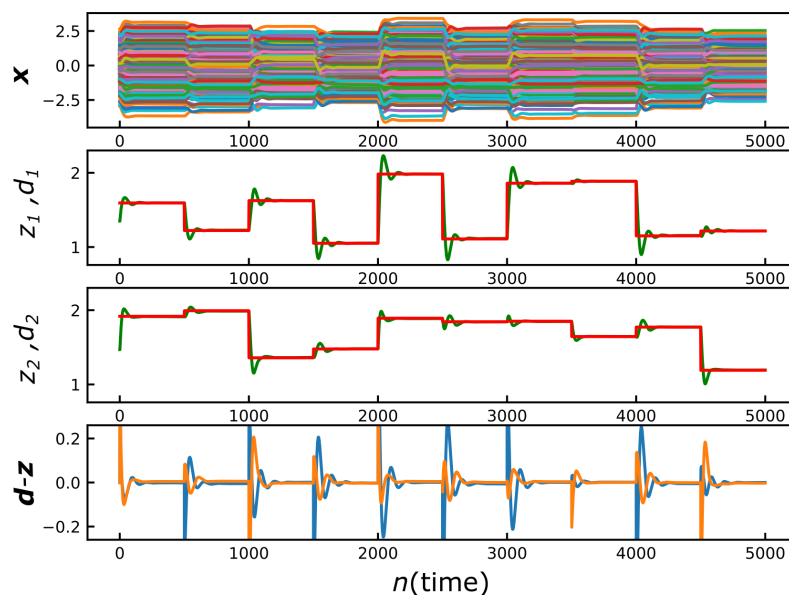


図 3.15 ネットワーク B のテスト入力 A に対する応答。2,3 行目の赤いプロットは感覚入力  $d(t)$ ，2,3 行目の緑のプロットは予測  $z(t)$  を表す。

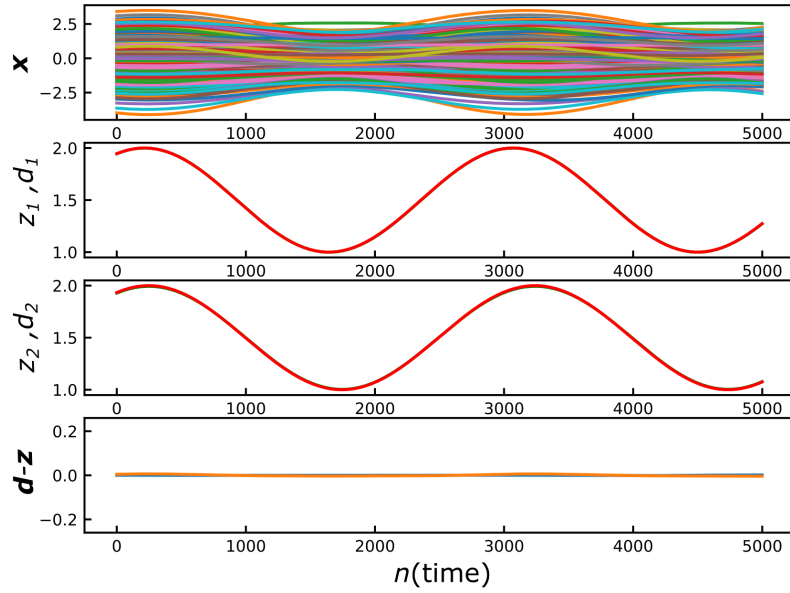


図 3.16 ネットワーク B のテスト入力 B に対する応答. 2,3 行目の赤いプロットは感覚入力  $d(t)$ , 2,3 行目の緑のプロットは予測  $z(t)$  を表す.

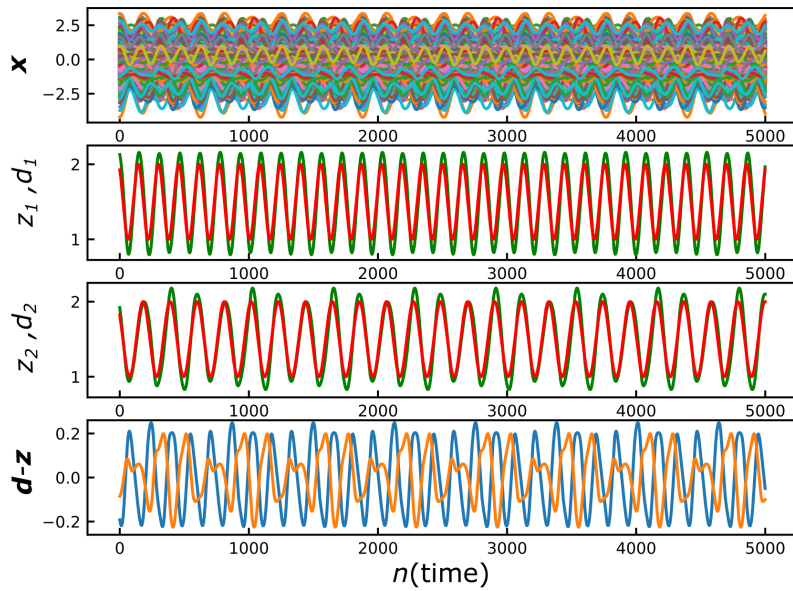


図 3.17 ネットワーク B のテスト入力 C に対する応答. 2,3 行目の赤いプロットは感覚入力  $d(t)$ , 2,3 行目の緑のプロットは予測  $z(t)$  を表す.

### 3.5.2 非線形ダイナミクス解析

前節で示したネットワーク B の挙動の背後にあるメカニズムを、非線形ダイナミクス解析によって明らかにする。基本的な流れはネットワーク A の時の解析に準じる。

まず、テスト入力 A の各試行  $i$  の最後における固定点  $\bar{x}^i (i = 1, 2, \dots)$  の位相空間における配置、および 1 試行における  $x$  の軌道の典型例を、図 3.18 に示す。この図より、ネットワーク A と同様に、ネットワーク B の位相空間上にも無数の固定点  $\bar{x}^i$  から構成される 2 次元多様体が存在することが分かる。また、 $x$  の軌道が試行の始めに 2 次元多様体上から引き離され、その後 2 次元多様体上に吸引されていく様子も確認できる。

次に、各固定点  $\bar{x}^i$  における式 (3.20) で定義される疑似エネルギー関数  $q(\bar{x}^i)$  の値を視覚化すると、図 3.19 のようになった。図 3.11 と比較すると、ネットワーク B の場合はほぼ全ての  $q(\bar{x}^i)$  の値が極めて小さく、これらの  $\bar{x}^i$  はネットワーク A の場合よりもより完全な固定点とみなせることが分かる。

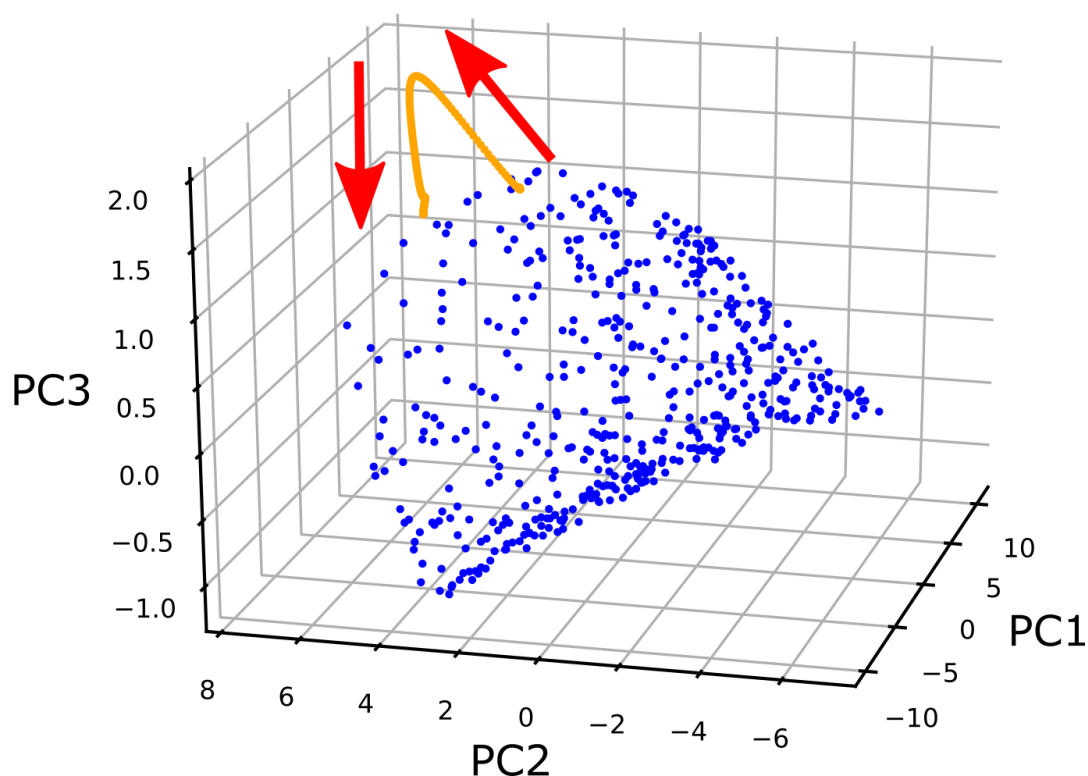


図 3.18 ネットワーク B の位相空間における固定点の位置 (青) と 1 試行あたりの軌道 (黄色)。

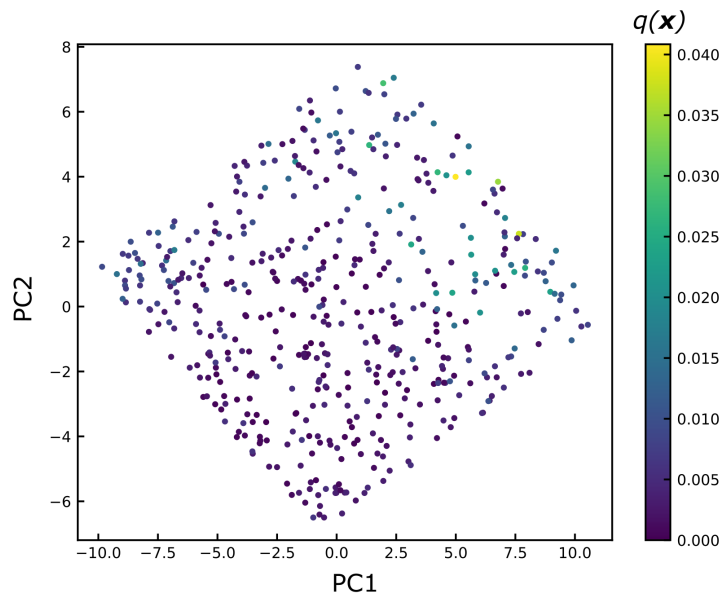


図 3.19 ネットワーク B の 2 次元多様体上における疑似エネルギー  $q(\mathbf{x})$  の値.

次に、この 2 次元多様体の安定性を調べるため、その上に分布する固定点  $\bar{\mathbf{x}}$  におけるヤコビ行列 (式 (3.21)) の固有値を求めた。結果、多くの固定点において、固有値の分布は図 3.20 のような形状であった。図 3.20 では、2 組の複素共役な固有値が実軸上にあり、残りの固有値の実部はすべて負であることが確認できる。これは、これらの固定点が  $N - 2$  方向に対して吸引的であり、残りの 2 方向に対してはセンターとして振舞うことを示す。

さらに、各固定点における固有値の最大実部の値を視覚化すると、図 3.21 のようになった。この図より、2 次元多様体上の固定点のほとんどが同様に振舞うと予想され、この 2 次元多様体は 1 つの中心多様体とみなせる。すなわち、この中心多様体上を構成する固定点はほぼ完全に連続的に並んでおり、中心多様体外の方向については吸引的で、中心多様体上の方向についてはセンターとして振舞う。従って、この中心多様体上でのフローの大きさはほぼ 0 となっている。これがネットワーク A の持つ 2 次元多様体アトラクターとの大きな違いであり、ネットワーク A に比べてテスト入力 A に対する予測誤差が極めて小さくなる理由である。

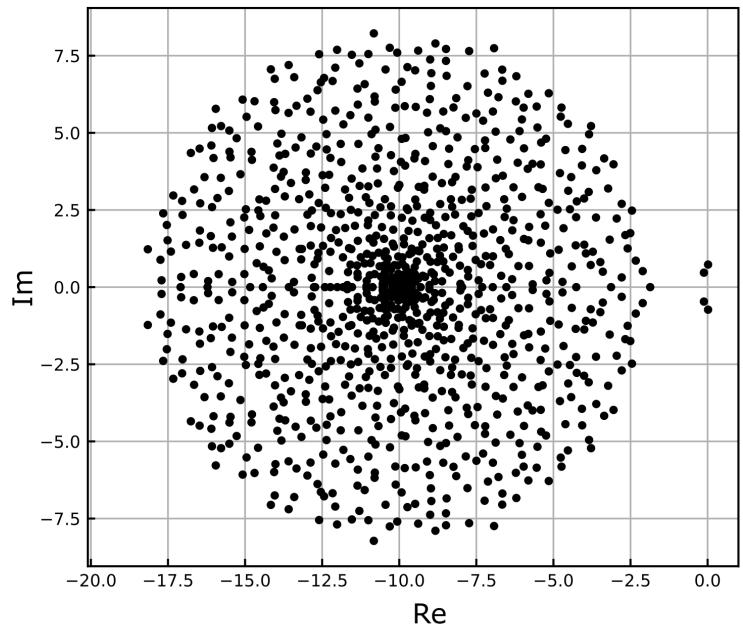


図 3.20 ネットワーク B の 2 次元多様体上の固定点のヤコビ行列における固有値スペクトラムの典型例.

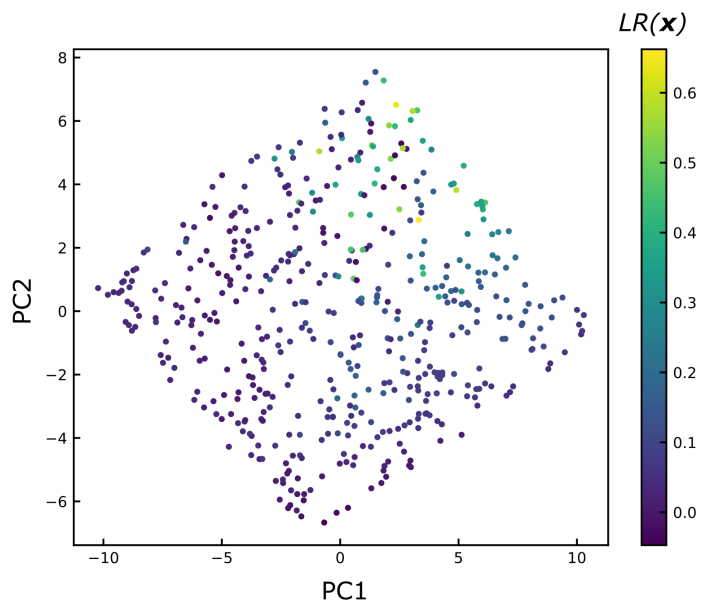


図 3.21 ネットワーク B の 2 次元多様体上の固定点  $\bar{x}$  の位置とヤコビ行列の固有値の最大実部  $LR(\bar{x})$  の関係.



テスト入力 B, テスト入力 C に対する応答についても, ネットワーク A の時と同様に, ネットワーク B はこの中心多様体を用いて予測を行っていると考えられる.

## 3.6 ネットワーク C : 急速に連続変化する教師信号

### 3.6.1 性能

テスト入力 A に対するネットワーク C の応答を図 3.22 に示す. この図が示す通り, ネットワーク C によるテスト入力 A の予測はほとんどの試行で失敗している.

次に, テスト入力 B に対するネットワーク C の応答を図 3.23 に示す. この図が示す通り, ネットワーク C はテスト入力 B を 0.2 以内の予測誤差で予測できていることが分かる. しかしこれはネットワーク A, ネットワーク B のテスト入力 B に対する予測精度には大きく劣る.

最後に, テスト入力 C に対するネットワーク C の応答を図 3.24 に示す. この図が示す通り, ネットワーク C はテスト入力 C を 0.2 以内の予測誤差で予測できていることが分かる. すなわち, ネットワーク B と同程度の予測精度であり, ネットワーク A には劣る.

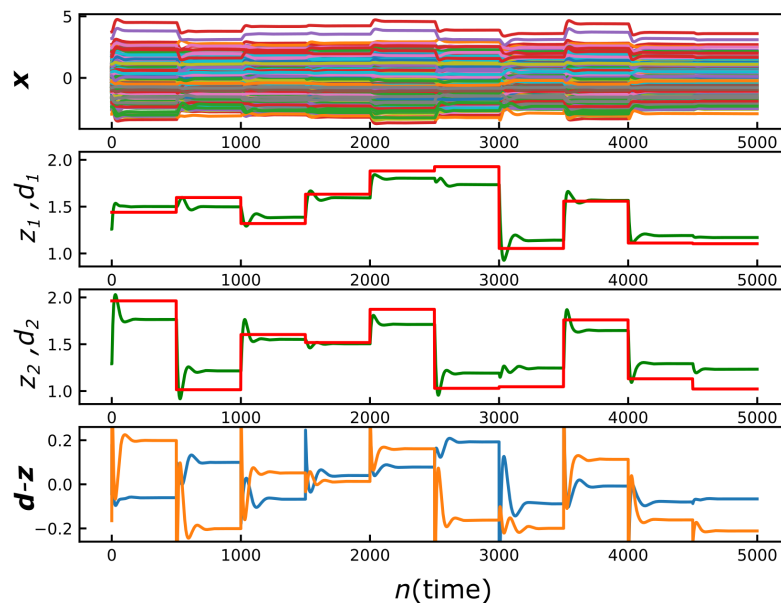


図 3.22 ネットワーク C のテスト入力 A に対する応答. 2,3 行目の赤いプロットは感覚入力  $d(t)$ , 2,3 行目の緑のプロットは予測  $z(t)$  を表す.

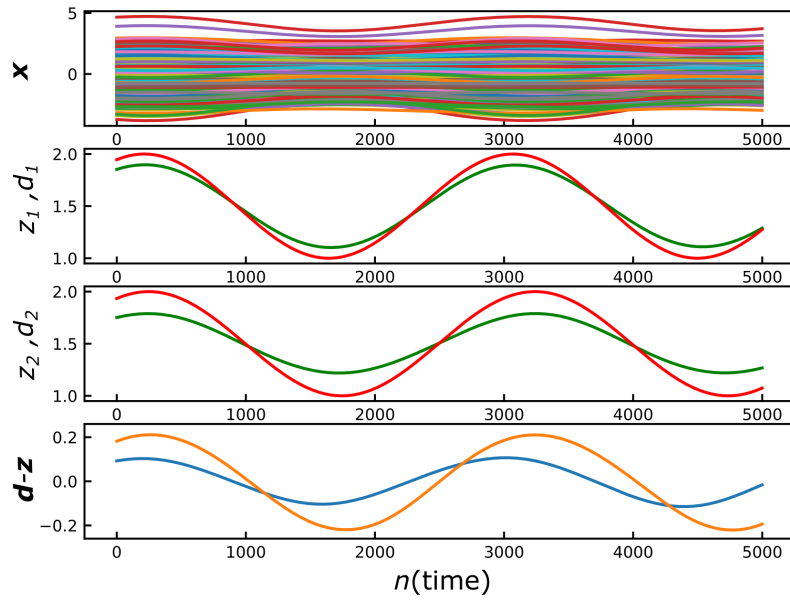


図 3.23 ネットワーク C のテスト入力 B に対する応答. 2,3 行目の赤いプロットは感覚入力  $d(t)$ , 2,3 行目の緑のプロットは予測  $z(t)$  を表す.

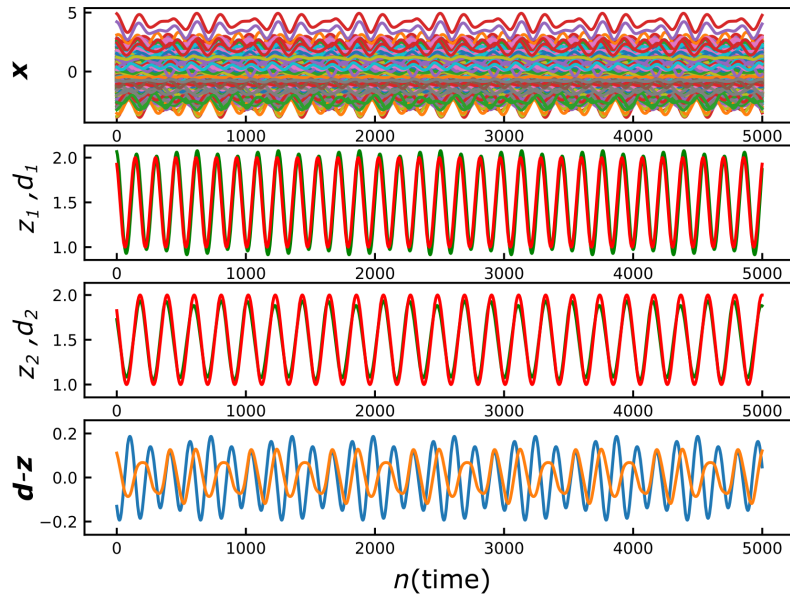


図 3.24 ネットワーク C のテスト入力 C に対する応答. 2,3 行目の赤いプロットは感覚入力  $d(t)$ , 2,3 行目の緑のプロットは予測  $z(t)$  を表す.

### 3.6.2 解析

図 3.22 の結果から、ネットワーク A やネットワーク B の場合とは異なり、ネットワーク C は固定点から成る 2 次元多様体を形成していないことが分かる。このため、未学習の時系列に対する予測がいずれも上手くいっていない。

ここで、訓練時に用いた教師信号をテスト時の感覚入力として与え、途中で予測誤差  $d - z$  のリザーバーへのフィードバックを切ってしまう場合を考える。この結果、ネットワーク C は図 3.25 のような応答を示した。この図を見ると、ネットワーク C は学習済みの時系列に対してはほぼ完全な予測が行えており、また予測誤差のフィードバックを切ると少しずつ予測誤差が増大していくことが分かる。これより、ネットワーク C は、予測誤差入力なしでも学習済みの時系列  $d(t)$  をほぼ正確に出力するための力学系を備えており、予測誤差入力のあるときはそのわずかな誤差を補正しながら学習済みの時系列に対して正確な行うことができる、と考えられる。

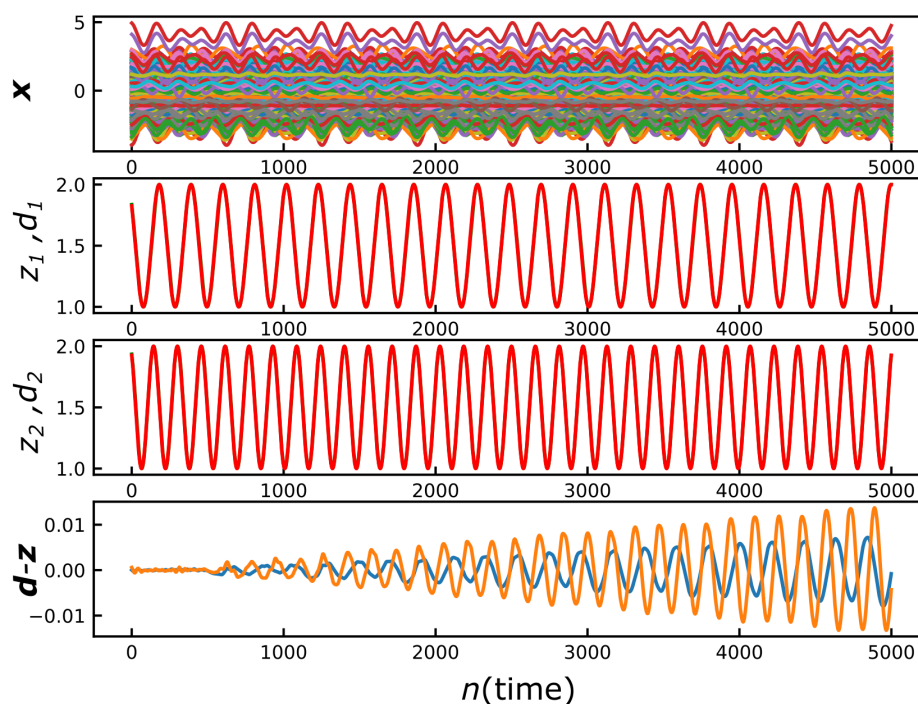


図 3.25 ネットワーク C の学習済み時系列に対する応答 ( $n = 1, \dots, 500$ ) と、その後予測誤差入力を切った場合の自発的な応答 ( $n = 501, \dots, 5000$ )。2,3 行目の赤いプロットは感覚入力  $d(t)$ ，2,3 行目の緑のプロットは予測  $z(t)$  を表す。

## 第 4 章

# 文脈依存型 PCRC モデル

本章では本研究の結果として，修正 PCRC モデルに外部モジュールから送られる定常的なコンテキスト信号を組み込んだ，文脈依存型 PCRC モデルの実装を説明する．また，不連続変化する教師信号を用いたネットワーク（前章のネットワーク A と同様）について予測性能を示し，コンテキストに応じた処理の切り替えのメカニズムを解析する．さらに，高次元の感覚入力として画像データを用いた場合の文脈依存型 PCRC モデルの挙動を調べる．

### 4.1 コンテキストとは

ここではコンテキストとは，状況，目標，関連する過去の経験，等の情報を指す．コンテキストの内的知覚生成への影響は強く，物理的には全く同じ感覚入力に対しても，コンテキストによって全く異なる内的知覚が形成されることが知られている．視覚におけるその典型例として，同じ手書き文字であっても文脈に応じて「数字の 0」に見えたり「アルファベットの O」に見えたりすることが挙げられる．

コンテキスト信号をリカレントニューラルネットワークに組み込んだ先行研究として，Mante らのモデル [19] が代表的である．Mante らは学習後のネットワークを非線形動力学の観点から解析し，与えられるコンテキスト信号に応じて異なるアトラクター地形が構成され，同じ入力に対しても異なる情報を選択的に抽出できることを確認した．また彼らは，マカクザルの神経行動学実験データとこのニューラルネットワークモデルの挙動が合致することを確認した．

本章では，前章で解析した修正 PCRC モデルにコンテキスト信号を組み込むことを試みる．その際，Mante らのモデルに倣い，コンテキスト信号を one-hot ベクトルとしてネットワークに提示する．また，この文脈依存型 PCRC モデルの性能を示した後，非線形動力学の観点から学習後のネットワークがコンテキストに応じて予測を切り替える仕組みを明らかにする．

## 4.2 ネットワーク構造, ダイナミクス

文脈依存型 PCRC モデルでは, 式 (3.1) に外部モジュールから送られるコンテキスト信号に対応する項を加えることで得られる, 次のダイナミクスを用いる:

$$\tau \dot{\mathbf{x}} = -\mathbf{x} + W^{REC} \mathbf{y} + W^{FB} \mathbf{z} + W^{IN} (\mathbf{d} - \mathbf{z}) + W^{CON} \mathbf{c}. \quad (4.1)$$

ここで,  $L \in \mathbb{N}$  をコンテキスト信号の次元をとし,  $\mathbf{c}(t) \in \mathbb{R}^L$  はコンテキスト信号を表すとする. また,  $W^{CON} \in \mathbb{R}^{N \times L}$  は外部モジュールからのシナプス強度を表す結合係数行列であり, その各要素は一様分布  $\mathcal{U}[-1, 1]$  から i.i.d. にサンプリングされる. また, リザーバ計算を用いるため  $W^{CON}$  は学習中固定される. その他の設定は前章の修正 PCRC モデルに準じる.  $W^{OUT}$  の学習にも, 前章と同様 FORCE アルゴリズム [17] を用いる. 本モデルのネットワーク構造を図 4.1 に示す.

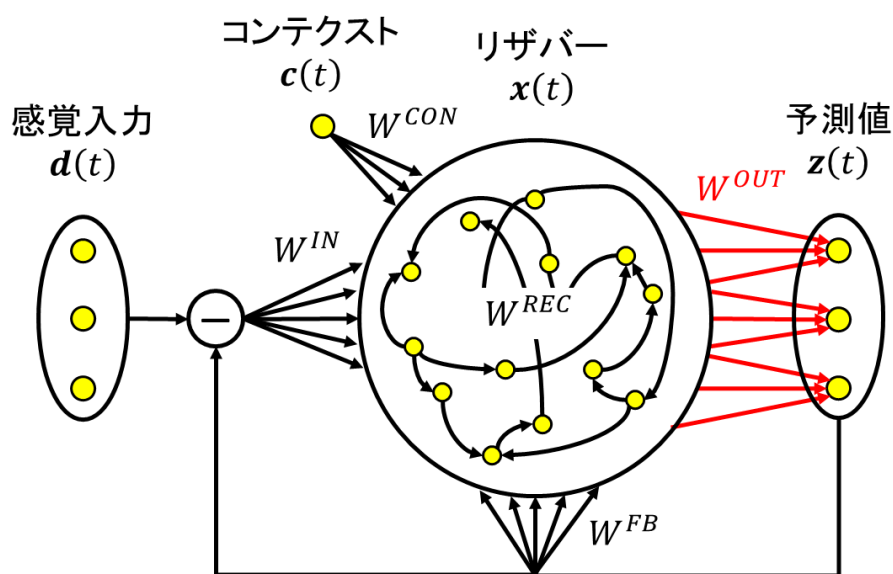


図 4.1 文脈依存型 PCRC モデルのネットワーク構造の概念図. 学習中は赤い結合  $W^{OUT}$  のみが更新される.

### 4.3 感覚入力が低次元の場合

本節を通して、数値シミュレーションでは簡単のため  $M = 4, L = 2$  を用いる。また、その他のパラメータ設定は前章のものに従う。

#### 4.3.1 訓練時

訓練時は、 $N_D = 2000$  を試行回数とし、 $i$  番目の試行 ( $i = 1, \dots, N_D$ ) において定ベクトル  $\mathbf{d}^i$  を感覚入力および教師信号としてネットワークに与える。同時に、 $1, \dots, \frac{N_D}{2}$  番目の試行ではコンテキスト信号として  $\mathbf{c}^0 := (0, 1)^T$  を、 $\frac{N_D}{2} + 1, \dots, N_D$  番目の試行ではコンテキスト信号として  $\mathbf{c}^1 := (1, 0)^T$  を、それぞれ与える。各感覚入力  $\mathbf{d}^i$  は  $0.2[\text{sec}]$  の間提示され、その値は次のように決められる：

$$\mathbf{d}^i = \left( d_1^i, \frac{1}{d_1^i}, d_2^i, \frac{1}{d_2^i} \right)^T \quad \text{if } \mathbf{c} = \mathbf{c}^0, \quad (4.2)$$

$$\mathbf{d}^i = \left( d_1^i, d_2^i, \frac{d_2^i}{2}, \frac{d_1^i}{2} \right)^T \quad \text{if } \mathbf{c} = \mathbf{c}^1. \quad (4.3)$$

但し、 $d_1^i, d_2^i$  は一様分布  $\mathcal{U}[1, 2]$  から i.i.d. にサンプリングされるとする。

この場合、 $\mathbf{c} = \mathbf{c}^0$  および  $\mathbf{c} = \mathbf{c}^1$  のそれぞれのコンテキスト下では、感覚入力の実質的な次元は  $M/2$  とみなせる。従って、前章のネットワーク A のように 1 つの  $M$  次元の多様体アトラクターを用意するよりも、各コンテキストにおいて  $M/2$  次元の異なる多様体アトラクターを用意することが自然な戦略だと考えられる。以下ではこれを実証していく。

#### 4.3.2 テスト時

テスト時には  $W^{OUT}$  は訓練後の状態で固定し、固定点への収束を確認するため各試行の長さは  $1.0[\text{sec}]$  とした。その他の設定は訓練時に準じた。

テスト時のネットワークの応答を図 4.2, 図 4.3 に示す。これらの図から分かる通り、学習後のネットワークによる予測はほとんど完全に成功している。また各コンテキストにおいて、前章のネットワーク A と同様に、試行の最初はネットワークはパルス状の予測誤差  $\mathbf{d}(t) - \mathbf{z}(t)$  によって固定点から引き離され、試行の最後はまた他の固定点に落ち着く様子が確認できる。

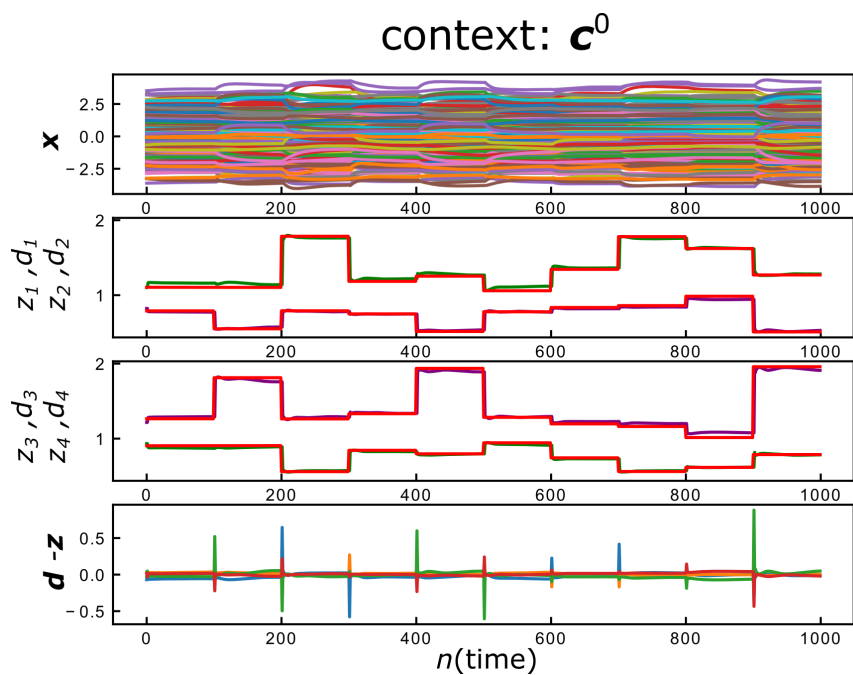


図 4.2 文脈依存型 PCRC モデルのテスト時における応答 (コンテキスト  $\mathbf{c} = \mathbf{c}^0$  下). 2,3 行目の赤いプロットは感覚入力  $\mathbf{d}(t)$ , 2,3 行目の緑と紫のプロットは予測  $\mathbf{z}(t)$  を表す.

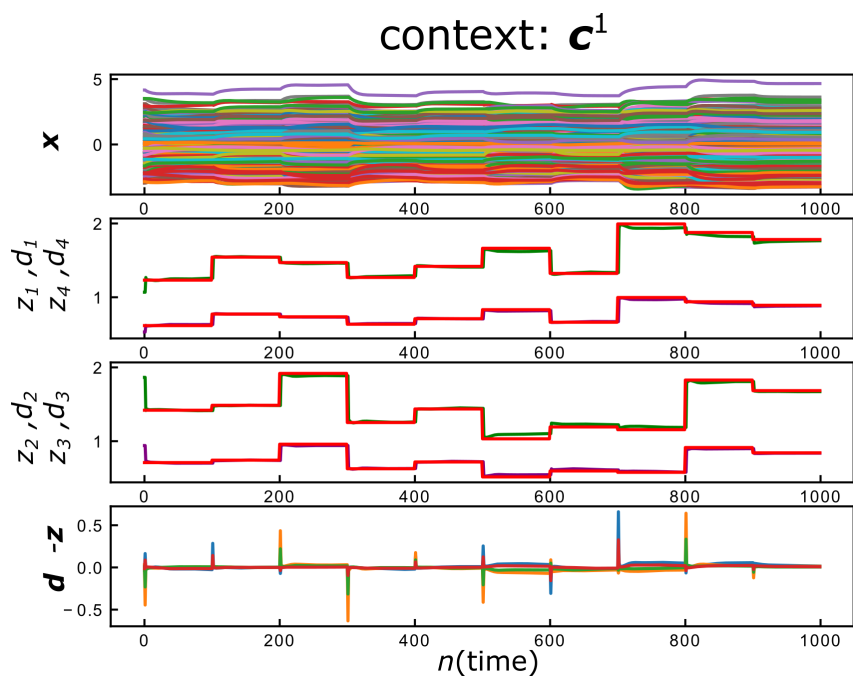


図 4.3 文脈依存型 PCRC モデルのテスト時における応答 (コンテキスト  $\mathbf{c} = \mathbf{c}^1$  下). 2,3 行目の赤いプロットは感覚入力  $\mathbf{d}(t)$ , 2,3 行目の緑と紫のプロットは予測  $\mathbf{z}(t)$  を表す.

また、各試行の最後における固定点  $\bar{\mathbf{x}}^i (i = 1, 2, \dots)$  の位相空間上での位置を PCA を用いて視覚化すると、図 4.4 のようになった。この図を見ると、各コンテキストにおいて、異なる 2 次元多様体上に固定点群が形成されている様子が分かる。これらの 2 次元多様体の安定性の解析はここでは省略するが、各コンテキストにおいて、前章のネットワーク A と同様にこれらの多様体がアトラクターとして振舞うことで予測を可能にしていると考えられる。

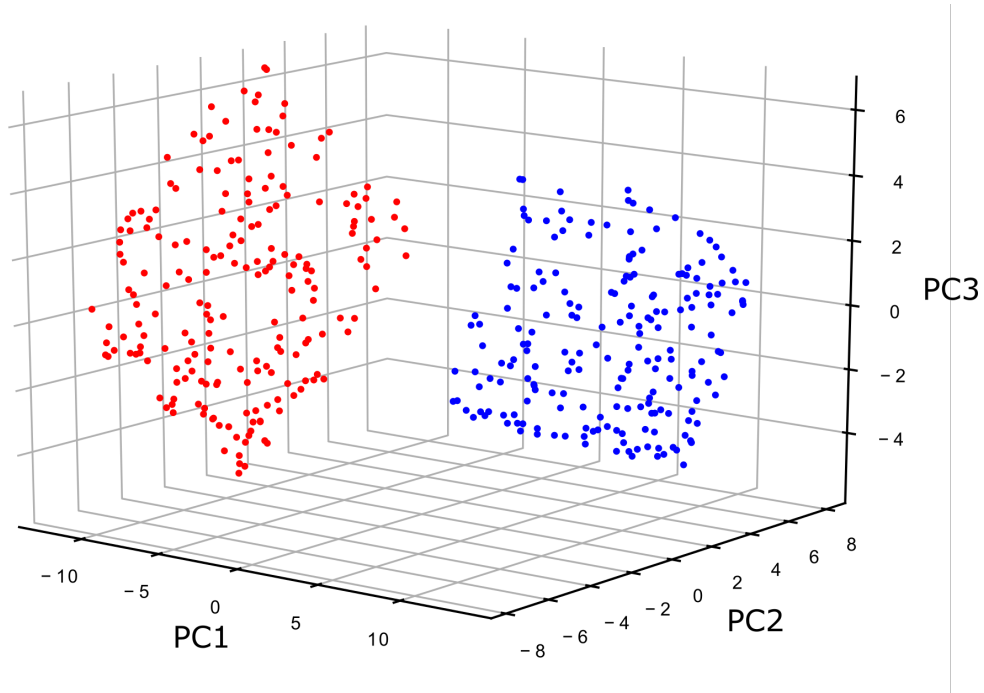


図 4.4 文脈依存型 PCRC モデルの位相空間における各コンテキスト下での固定点の配置. 青 :  $\mathbf{c} = \mathbf{c}^0$ , 赤 :  $\mathbf{c} = \mathbf{c}^1$ .

### 4.3.3 コンテキストミスマッチ

次に、受け取ったコンテキスト信号  $\mathbf{c}$  と実際の感覚入力  $\mathbf{d}$  のタイプが訓練時と一致しない、コンテキストミスマッチの場合を考える。すなわち、テスト時にコンテキスト  $\mathbf{c}$  と感覚入力  $\mathbf{d}$  をそれぞれ、

$$\mathbf{d}^i = (d_1^i, \frac{1}{d_1^i}, d_2^i, \frac{1}{d_2^i})^T \text{ if } \mathbf{c} = \mathbf{c}^1, \quad (4.4)$$

$$\mathbf{d}^i = (d_1^i, d_2^i, \frac{d_2^i}{2}, \frac{d_1^i}{2})^T \text{ if } \mathbf{c} = \mathbf{c}^0, \quad (4.5)$$

として与えた。後に示すように、固定点に収束しないケースも多くあるため、各試行の長さは 5.0[sec] とした。



コンテキストミスマッチ時のネットワークの応答を図 4.6, 図 4.5 に示す. これらの図が示す通り, コンテキストミスマッチ時には, ネットワークの予測は全く失敗してしまう. また, 固定点への収束速度が非常に遅かったり, そもそも固定点に収束しなかったりと, 多様体アトラクターでは説明できないダイナミクスが働いていることが分かる. これらの結果から, 予測を行うための多様体アトラクターはあくまでコンテキストに対応したタイプの入力に対してのみ形成されていることが分かる. すなわち, 図 4.4 の青プロットの多様体は  $\mathbf{c} = \mathbf{c}^1$  下では存在せず, 赤プロットの多様体も  $\mathbf{c} = \mathbf{c}^0$  下では存在しない. 従って, 文脈依存型 PCRC モデルによる予測は, あくまでコンテキストミスマッチが起きない時のみ正常に行われるということが分かる.

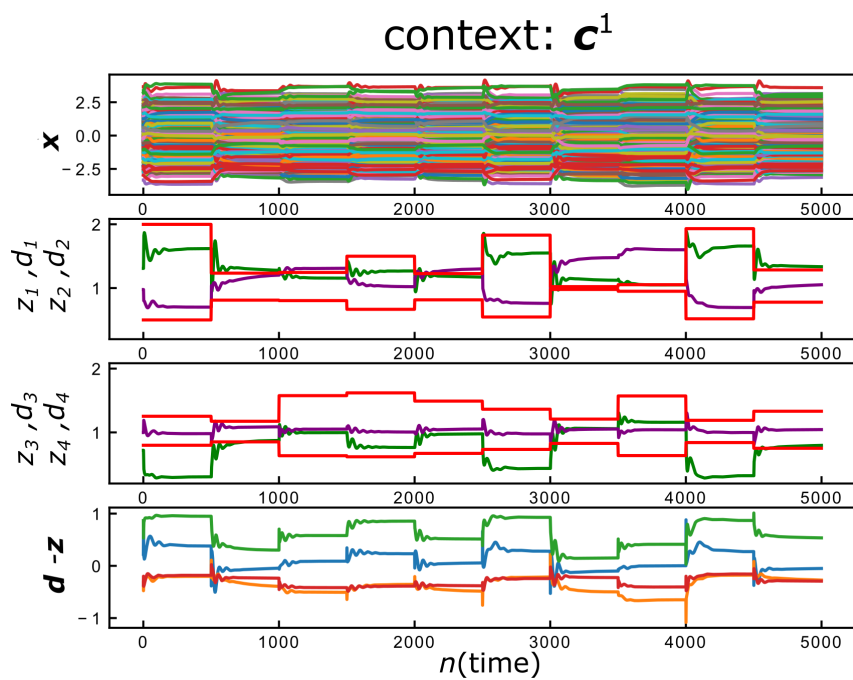


図 4.5 文脈依存型 PCRC モデルのコンテキストミスマッチ時における応答 ( $c = c^1$  下). 2,3 行目の赤いプロットは感覚入力  $d(t)$ , 2,3 行目の緑と紫のプロットは予測  $z(t)$  を表す.

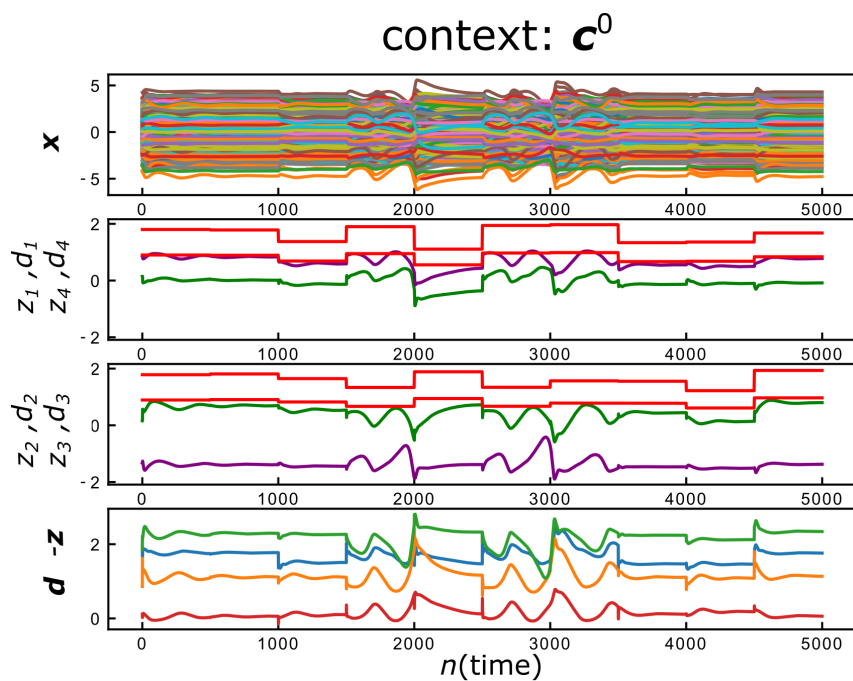


図 4.6 文脈依存型 PCRC モデルのコンテキストミスマッチ時における応答 ( $c = c^0$  下). 2,3 行目の赤いプロットは感覚入力  $d(t)$ , 2,3 行目の緑と紫のプロットは予測  $z(t)$  を表す.

## 4.4 感覚入力が高次元の場合

文脈依存型 PCRC モデルが視覚系における内的知覚の生成を説明するモデルとして有力であることを示すために、本節では感覚入力に高次元の画像データを用いてシミュレーションを行う。具体的には、手書き数字認識のタスクで広く用いられる MNIST (Mixed National Institute of Standards and Technology) データセットを視覚入力として採用した。

本節を通して、数値シミュレーションでは簡単のため  $M = 20, L = 2$  を用いる。また、その他のパラメータ設定は前節に準じる。

### 4.4.1 訓練時

教師信号および感覚入力として用いる MNIST データは、ラベルが 0 または 1 のものに限る。また、MNIST データは元々 784 次元だが、ここでは事前処理として非負値行列分解 (NMF: Non-negative Matrix Factorization) を用いて、 $M = 20$  次元まで圧縮しておく。これはリザーブの次元  $N = 1000$  に対して入出力の次元  $M$  が大きいと予測がうまくいかないためである。

訓練時は、 $N_D = 4000$  回の試行を繰り返す。1, ...,  $\frac{N_D}{2}$  番目の試行ではラベル 0 の圧縮 MNIST データのうちから、 $\frac{N_D}{2} + 1, \dots, N_D$  番目の試行ではラベル 1 の圧縮 MNIST データのうちから、ランダムに 1 つ選び、感覚入力  $\mathbf{d}^i$  としてネットワークに与える。同時に、感覚入力  $\mathbf{d}^i$  のラベルが 0 ならばコンテキスト信号として  $\mathbf{c}^0 := (0, 1)^T$  を、ラベルが 1 ならばコンテキスト信号として  $\mathbf{c}^1 := (1, 0)^T$  を、それぞれ与える。各試行の長さは 0.2[sec] とする。

この場合、コンテキストは入力画像のカテゴリ、すなわち「数字の 0」か「数字の 1」かを表している。訓練を通してネットワークは、前節のネットワークのように、各コンテキストにおいてそれぞれのカテゴリに対応する多様体アトラクターを用意することで知覚を可能にすると予想される。

### 4.4.2 テスト時

テスト時には  $W^{OUT}$  は訓練後の状態で固定し、固定点への収束を確認するため各試行の長さは 5.0[sec] とした。また、ラベルが 0 または 1 の圧縮 MNIST データのうち、訓練時には選ばなかったものをテスト時の感覚入力として用いた。これにより、訓練によって未学習の MNIST データの知覚も可能になるかどうかを調べた。その他の設定は訓練時に準じた。

テスト時のネットワークの応答を図 4.7, 図 4.8 に示す。これらの図が示す通り、学習後のネットワークによる未学習の MNIST データの予測はほとんど成功している。予測誤

差  $d - z$  の値は感覚入力  $M = 2$  次元の場合のネットワーク (図 4.2, 図 4.3) に比べると大きくなってしまっているが, これは入力の次元が大きい割に学習に使うデータの数小さいこと起因すると考えられる. また, 各コンテキストにおいて, これまでと同様にパルス的な予測誤差  $d - z$  に駆動されてネットワークが固定点から固定点に移動している様子が確認できる.

さらに, 各試行の最後における固定点  $\bar{x}^i (i = 1, 2, \dots)$  の位相空間上での位置を PCA を用いて視覚化すると, 図 4.9 のようになった. この図を見ると, 各コンテキストにおいて, 位相空間に異なる固定点群が形成されている様子が分かる. これらの固定点群は入出力の次元  $M$  より低い次元の多様体上に分布していることが予想されるが, その正確な形状は 3 次元の PCA 空間においては視認できない. これらの多様体の安定性の解析はここでは省略するが, 図 4.7, 図 4.8 の結果から, 各コンテキストにおいてこれらの多様体がアトラクターとして振舞うことで予測を可能にしていると考えられる.

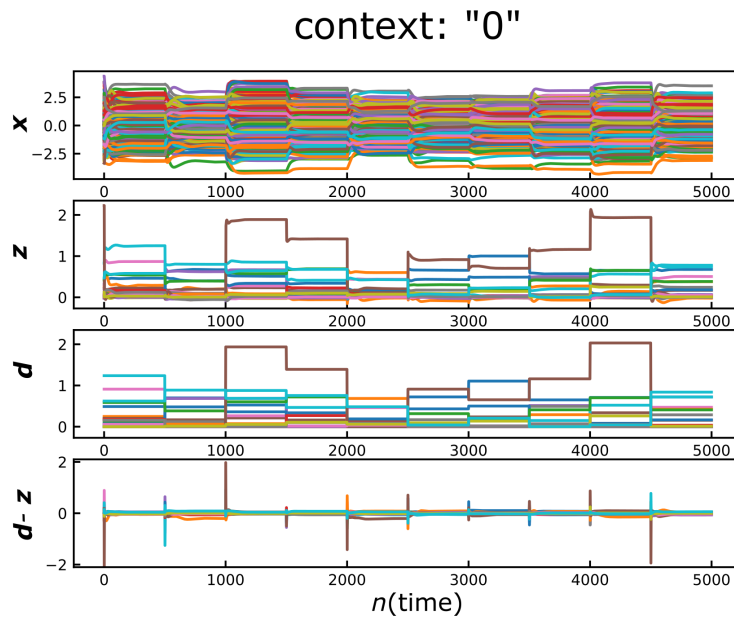


図 4.7 MNIST データを学習した文脈依存型 PCRC モデルのテスト時における応答 (コンテキスト:  $c = c^0$ , 感覚入力のラベル: “0”).

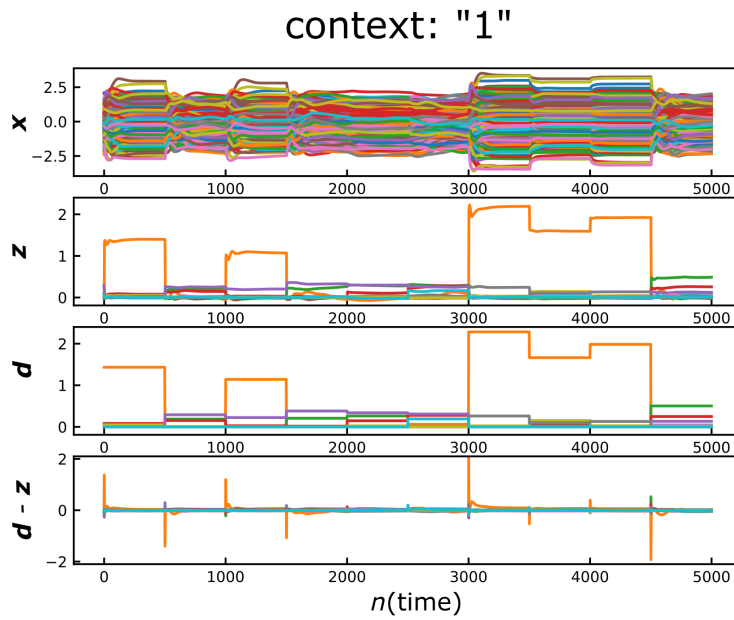


図 4.8 MNIST データを学習した文脈依存型 PCRC モデルのテスト時における応答 (コンテキスト:  $c = c^1$ , 感覚入力のラベル: “1”).

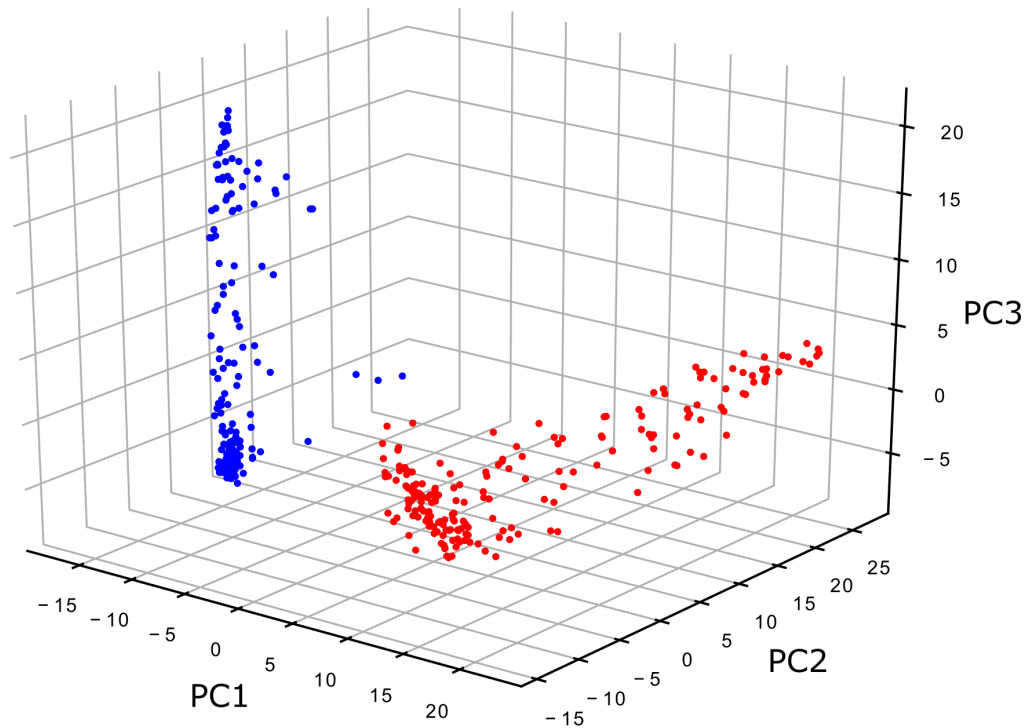


図 4.9 MNIST データを学習した文脈依存型 PCRC モデルの位相空間における各コンテキスト下での固定点の配置. 青 : ラベル 0 ( $c = c^0$ ), 赤 : ラベル 1 ( $c = c^1$ ).

#### 4.4.3 コンテキストミスマッチ

次に、受け取ったコンテキスト信号  $c$  と感覚入力  $d$  のラベルが訓練時と一致しない、コンテキストミスマッチの場合を考える。

テスト時に、コンテキスト  $c = c^1$  下でラベル 0 の感覚入力を、コンテキスト  $c = c^0$  下でラベル 1 の感覚入力  $d$  をそれぞれ与えると、ネットワークの応答は図 4.10, 図 4.11 のようになった。これらの図を見ると、予測誤差  $d - z$  はどの試行においても 0 から遠く、予測は成功とはいえない。しかし、前節のコンテキストミスマッチ時の結果 (図 4.5, 図 4.6) に比べると固定点に落ち着くことが多く、また  $d$  と  $z$  の形状を比較してみても、部分的に予測が成功していることが確認できる。これは入力が高次元の場合にのみ起こる現象であり、予測値の  $M$  個の成分のうち、コンテキスト信号  $c$  の寄与が大きい部分と小さい部分に分かれるためと考えられる。

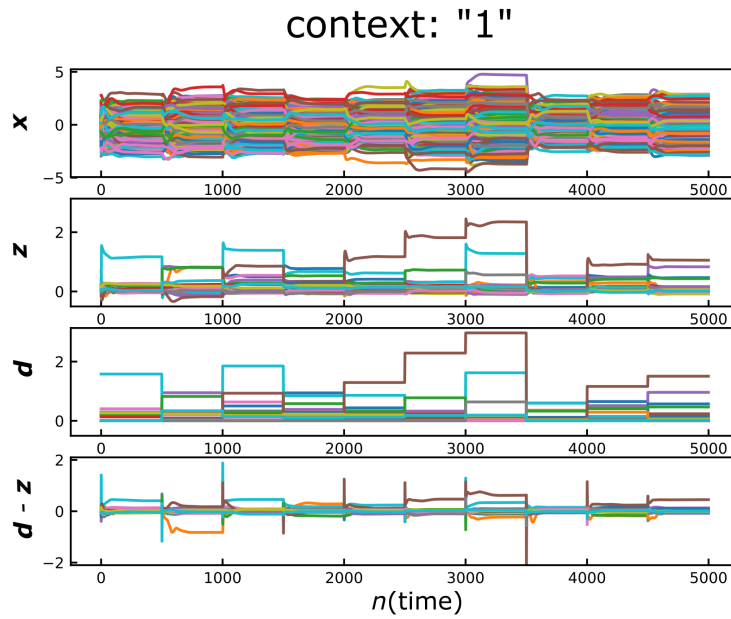


図 4.10 MNIST データを学習した文脈依存型 PCRC モデルのコンテキストミスマッチ時における応答 (コンテキスト:  $c = c^1$ , 感覚入力のラベル: “0”).

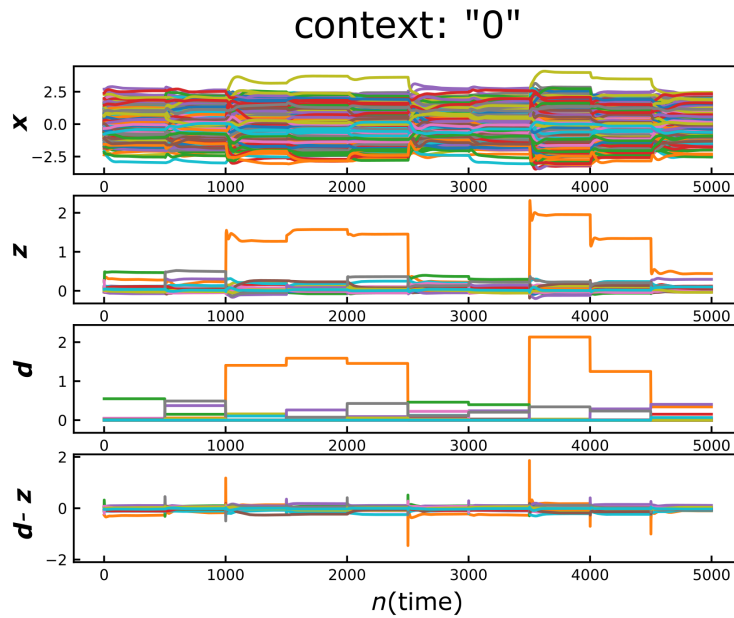


図 4.11 MNIST データを学習した文脈依存型 PCRC モデルのコンテキストミスマッチ時における応答 (コンテキスト:  $c = c^0$ , 感覚入力のラベル: “1”).

コンテキストミスマッチ時の固定点における予測値  $\mathbf{d}$  の特徴をさらに調べるため、NMF の逆変換を用いて  $\mathbf{d}$  を MNIST データの元の次元 (784 次元) に展開して可視化し、感覚入力  $\mathbf{d}$  の元となる MNIST データと比較した。可視化した  $\mathbf{d}, \mathbf{z}$  の比較の一部を図 4.12, 図 4.13 に示す。

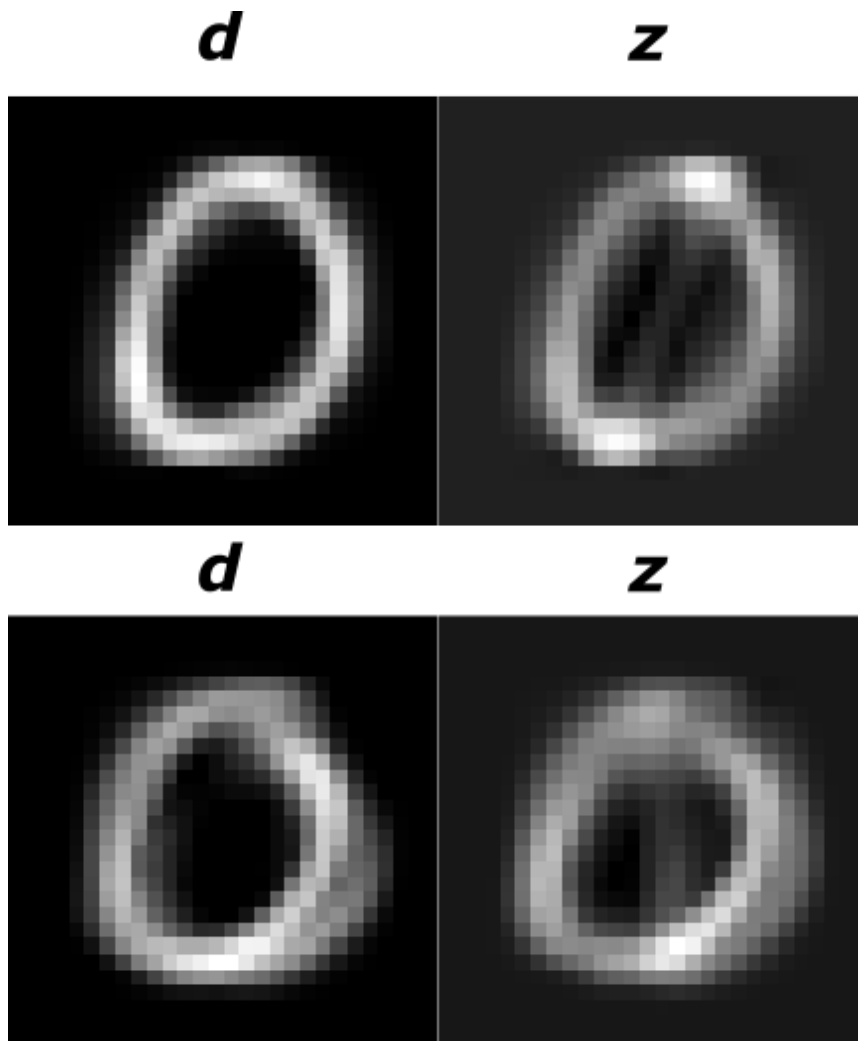


図 4.12 コンテキストミスマッチ時における予測  $\mathbf{z}$  と感覚入力  $\mathbf{d}$  の可視化による比較 (コンテキスト :  $\mathbf{c} = \mathbf{c}^1$ , 感覚入力のラベル : “0”)



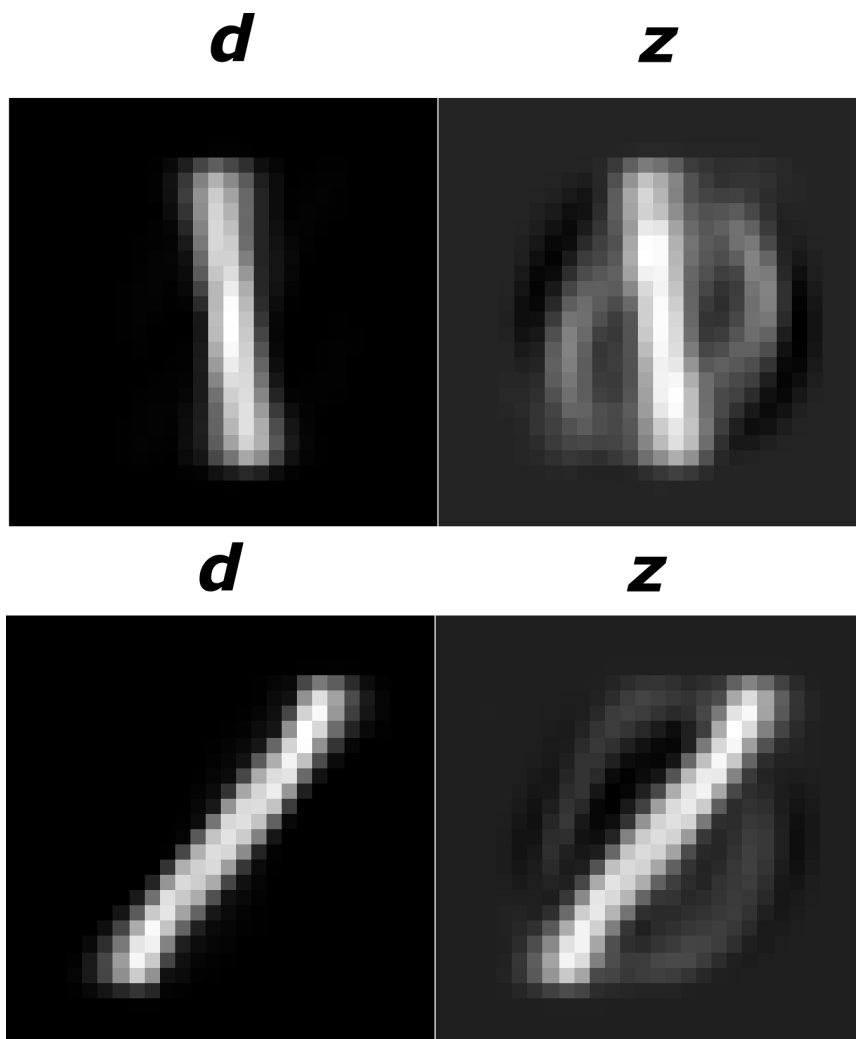


図 4.13 コンテキストミスマッチ時における予測  $z$  と感覚入力  $d$  の可視化による比較  
(コンテキスト:  $c = c^0$ , 感覚入力のラベル: “1”)

結果として、感覚入力が高次元の場合のコンテキストミスマッチ時の固定点における予測  $z$  は、次の 2 つの成分が重なったものであることが確認できた：

- 感覚入力  $d$  に対する予測.
- コンテキスト信号  $c$  の示すラベルに対応する予測.

また、今回の場合では後者の成分より前者の成分の方が強く現れていることも確認できた.

## 第 5 章

# 結論，考察

本章では，本研究で構築した修正 PCRC モデルと文脈依存型 PCRC モデルについて，その解析結果に対する考察を行う．

### 5.1 修正 PCRC モデル

3 章では，Fukino モデルの欠点を克服した修正 PCRC モデルを構築し，訓練時の教師信号が，不連続変化する場合（ネットワーク A），ゆっくりと連続変化する場合（ネットワーク B），急速に連続変化する場合（ネットワーク C），の 3 つのネットワークを用意し，それぞれの予測性能を調べた．3 章で示した性能の違いをまとめると次のようになる．

- 不連続変化する感覚入力（テスト入力 A）に対する応答：  
最終的な予測精度はネットワーク B が最も高いが，固定点への収束速度はネットワーク A が最も速かった．ネットワーク C は予測に失敗してしまった．
- ゆっくりと連続変化する感覚入力（テスト入力 B）に対する応答：  
予測精度はネットワーク B が最も高く，ネットワーク C が最も低かった．
- 急速に連続変化する感覚入力（テスト入力 C）に対する応答：  
予測精度はネットワーク A が最も高く，ネットワーク B，C は同程度に予測誤差があった．

さらに，各場合における訓練後のネットワークの予測メカニズムについて，非線形力学の観点から解析を行った．結果として，ネットワーク A，B は共に位相空間内に低次元の多様体アトラクターを形成することで未学習の感覚入力に対する予測を可能にしていることが分かった．一方，ネットワーク C は訓練時の教師信号そのものを再現する力学系を形成してしまい，多様体アトラクターの形成ができなかったため，学習済みの時系列に関しては完全な予測ができるが，未学習の時系列の予測はできなくなっていることが分かった．

また、ネットワーク A と B では多様体アトラクターに吸引される速度が異なり、その結果、テスト入力 A やテスト入力 C 等の急速な変化を伴う時系列に対する応答にも違いが見られた。これは、ネットワーク A は訓練時に感覚入力之急激に変化することで固定点間の移動も急速に起こることを要求されたが、ネットワーク B は訓練時に感覚入力が一歩ゆっくりとしか変化しないため急速な応答が必要とされなかったから、と考えられる。

以上のことから、訓練時に用いた教師信号の変化の時間スケールによって、訓練後のネットワークが獲得するダイナミクスの性質が大きく変わり、結果としてそのネットワークが予測を得意とする時系列の種類も変わる、ということが分かる。

## 5.2 文脈依存型 PCRC モデル

4 章では、修正 PCRC モデルに外部モジュールから送られる定常的なコンテキスト信号を組み込むことで、文脈依存型 PCRC モデルを構築し、感覚入力が低次元の場合と高次元の場合の 2 つのネットワークについて、予測性能とそのメカニズムを調べた。3 章のネットワーク A と同様に、訓練時は不連続変化する教師信号を用いた。

結果として、文脈依存型 PCRC モデルはコンテキストに応じて異なる多様体アトラクターを位相空間内に形成することで、コンテキストに応じた予測の切り替えを可能にしていることが分かった。

また、コンテキスト信号と感覚入力のタイプが訓練時の組み合わせとは異なる時 (コンテキストミスマッチ) にネットワークがどのような応答を示すかも調べた。結果として、コンテキストミスマッチ時においてネットワークの予測は失敗した。また、高次元の感覚入力を用いた場合に特有の現象として、感覚入力に応じた正しい予測値とコンテキスト信号に応じた誤った予測値の 2 つの成分が合成されて予測が形成されることが確認された。

ここで確認された誤った予測の特徴は、レヴィー小体型認知症における幻覚症状に見られる特徴 [20] と共通点を持つ。その幻覚症状では、目の前に存在する背景の上に、そこにいないはずの人の顔が見えることが報告されている。本モデルにおいてコンテキスト  $c = c^0$  を背景、コンテキスト  $c = c^1$  を人の顔と対応させると、現実にはそこに人がいなくても、人の顔に対応するコンテキスト信号が外部モジュールの誤作動によって送られてくると実際の背景に重なって人の顔が見える、というシナリオによる幻覚症状の説明が考えられる。

## 5.3 今後の課題

文脈依存型 PCRC モデルにおいて、コンテキスト信号は定常かつ外部モジュールから送られるものと仮定していた。このモデルの拡張として、コンテキスト信号が時間変化する場合や、コンテキスト信号自体を生成する高次のリザーバーの存在を仮定する場合が考えられる。さらに、その高次のリザーバーと低次の文脈依存型 PCRC モデルとの相互作用を

考慮することで，階層型 PCRC モデルを構築することが可能である．この階層型 PCRC モデルの挙動やメカニズムを調べ，従来の階層型予測符号化モデル [11][12][13] との比較を行うことが今後の課題である．

# 謝辞

本研究の進行に際し、指導教員の河野崇教授に終始適切なお助言と丁寧なご指導を頂いたことをここに深く御礼申し上げます。また、共同研究者として有益な議論を交わして下さった合原一幸教授、香取勇一准教授にも深く感謝致します。共に励まし合い、切磋琢磨することができた同期の学生に深く感謝致します。最後に、温かく見守り研究に集中できるよう支えてくれた家族に心から感謝します。

## 参考文献

- [1] S. Haykin and Z. Chen. The cocktail party problem. *Neural Computation*, 17(9):1875–1902, 2005.
- [2] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proc. R. Soc. Lond. B*, 204(1156):301–328, 1979.
- [3] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho. The mismatch negativity (mmn) in basic research of central auditory processing: a review. *Clinical Neurophysiology*, 118(12):2544–2590, 2007.
- [4] G. Stefanics, J. Kremláček, and I. Czigler. Visual mismatch negativity: a predictive coding view. *Frontiers in Human Neuroscience*, 8:666, 2014.
- [5] M. Garrido, J. M. Kilner, K. Stephan, and K. Friston. The mismatch negativity: a review of underlying mechanisms. *Clinical Neurophysiology*, 120(3):453–463, 2009.
- [6] C. Wacongne, J. P. Changeux, and S. Dehaene. A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, 32(11):3665–3678, 2012.
- [7] T. Nagai, M. Tada, K. Kirihara, T. Araki, S. Jinde, and K. Kasai. Mismatch negativity as a “translatable” brain marker toward early intervention for psychosis: a review. *Frontiers in Psychiatry*, 4:115, 2013.
- [8] D. J. Felleman and D. E. Van. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1):1–47, 1991.
- [9] S. M. Sherman and R. W. Guillery. On the actions that one nerve cell can have on another: distinguishing “drivers” from “modulators” . *Proceedings of the National Academy of Sciences*, 95(12):7121–7126, 1998.
- [10] A. M. Rosier, L. Arckens, G. A. Orban, and F. Vandesande. Laminar distribution of nmda receptors in cat and monkey visual cortex visualized by [3h]-mk-801 binding. *Journal of Comparative Neurology*, 335(3):369–380, 1993.
- [11] R. P. Rao and D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci*,

- 2(1):79–87, 1999.
- [12] K. Friston and S. Kiebel. Predictive coding under the free-energy principle. *Philos. T. Roy. Soc. B.*, 364(1521):1211–1221, 2009.
  - [13] K. Ara and Y. Katori. Hierarchical network model of auditory information processing using dynamical predictive coding. In *23rd Int. Symp. Artificial Life and Robotics*, pages 41–46, 2018.
  - [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
  - [15] M. Fukino, Y. Katori, and K. Aihara. A computational model for pitch pattern perception with the echo state network. In *2016 Int. Symp. Nonlinear Theory and Its Application*, pages 271–274, 2016.
  - [16] M. Lukoševičius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.
  - [17] D. Sussillo and L. F. Abbott. Generating coherent patterns of activity from chaotic neural networks. *Neuron*, 63(4):544–557, 2009.
  - [18] D. Sussillo and O. Barak. Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Computation*, 25(3):626–649, 2013.
  - [19] V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78, 2013.
  - [20] D. Collerton, J. P. Taylor, I. Tsuda, H. Fujii, S. Nara, K. Aihara, and Y. Katori. How can we see things that are not there?: current insights into complex visual hallucinations. *Journal of Consciousness Studies*, 23(7-8):195–227, 2016.

## 発表文献

- [1] H. Tamura, T. Kohno, and K. Aihara. ReLU chaotic neuron model for deep chaotic neural networks. *生産研究*, 70:163-165, 2018.
- [2] H. Tamura and K. Aihara. Hebbian learning in a chaotic neural network in the chaotic itinerancy state. The 1st Retreat International Research Center for Neurointelligence, Yokohama(Japan), 2018.
- [3] N. So, T. Hayami, H. Tamura, H. Mizutani, and H. Yamakawa. Neocortical Functional Hierarchy Estimated from Connectomic Morphology in the Mouse Brain. *Biologically Inspired Cognitive Architectures 2018*, Prague(Czech), 2018.
- [4] H. Tamura, Y. Katori, and K. Aihara. Possible Mechanism of Internal Visual Perception: Context-dependent Processing by Predictive Coding and Reservoir Computing Network. *The 2019 International Conference on Artificial Life and Robotics*, Beppu(Japan), 2019.