

リカレントニューラルネットワークによる楽音合成

大矢 健一

国立長野工業高等専門学校 電子情報工学科

E-mail: ohya@ei.nagano-nct.ac.jp

楽音合成に関する研究は古く、これまでにさまざまなモデルが提示されている。本稿では、新たにリカレントニューラルネットワークによる楽音合成モデルを提示する。このモデルは非線型連立微分方程式で記述されるシステムであり、一般に解析的には解けないために数値計算により解かれる。このモデルは、ある条件においては振動する解を得ることができ、そのふるまいは非線型ダイナミクスの特徴であるゆらぎやカオスといった特徴を持つ。また、このモデルは学習則により非線型ダイナミクスを学習することも可能である。

A Sound Synthesis by Recurrent Neural Network

Ken'ichi OHYA

Dept. of Electronics and Computer Science, Nagano National College of Technology

716 Tokuma, Nagano City, Nagano 381, JAPAN

E-mail: ohya@ei.nagano-nct.ac.jp

Study of sound synthesis has a long history in computer music and many models have been presented. In this paper, a new sound synthesis model using recurrent neural network is shown. This model is a system of nonlinear simultaneous differential equations. They are not solved analytically but solved numerically. The solution of the model oscillates under some conditions of parameters and its behavior has features of fluctuation or chaos, which are well known as features of nonlinear dynamics. Moreover, this model can learn nonlinear dynamics by the learning rule.

1 はじめに

楽音合成の技術にはさまざまなものがあるが、ここに新たに、リカレントニューラルネットワークによる楽音合成 (Recurrent Neural Network Synthesis. 以下、RNNS) を提示する。

リカレントニューラルネットワークは、数個のニューロンからなる単純なシステムであっても、リカレント結合の存在により複雑なふるまいを示す。この性質を用いて、リカレントニューラルネットワークをリズム認知モデルに応用した例を筆者はすでに報告している [Ohya, 1994a] [Ohya, 1994b]。

リカレントニューラルネットワークは、非線型ダイナミクスやカオスダイナミクスを出力できること

が知られており [Sato, 1990b] [Sato, 1993]、今回の研究ではこれを楽音合成に応用した。

RNNS は、楽音合成の手法としては非線型変換に属する。RNNS の特徴は以下のようにまとめることができる。

1. ニューロンのダイナミクスを楽音に用いる
2. FM (Frequency Modulation) 合成との類似
3. 数個のニューロンの場合にも豊かな波形
4. 多くのニューロンの場合には学習則を持つ

第1点であるが、ニューロン単一のモデルとして連続時間・連続値のモデルを採用する。これにより、ニューロンの出力値は時間的にも滑らかに変化する。

第2点は、FM合成における各オペレータとRNNSにおける各ニューロンとの対応関係による類似点を示している。第3点は、数個のニューロンの場合にも、連続時間・連続値のモデルによるリカレントニューラルネットワークのために、楽音として十分に豊かな波形を出力することを示している。第4点は、ニューロンの数が十分に多いときには、学習則により与えられた波形を学習可能であることを示す。このときの学習は、単純な層構造によるバックプロパゲーションではなく、時間的に変化するダイナミクスをシステムは学習することになる。よって、波形が元々含むゆらぎやカオスのな性質をも学習可能である。一般に、非線型変換による楽音合成では与えられた音を再合成することが難しいが、RNNSではそれが可能である。

2 ニューロンのモデル

2.1 単一ニューロン

ニューロン単一のモデルとして連続時間・連続値のモデルを採用する。本研究ではニューロンの出力値をそのまま音源の出力波形として用いるため、以上の条件は必須となる。

ニューロンのふるまいは以下の微分方程式で記述される。

$$\tau \frac{du}{dt} = -u + f\left(\sum_{j=1}^n W_j u_j\right) \quad (1)$$

ここに、

- $u(t)$ 時刻 t におけるニューロンの出力
- τ ニューロンの時定数
- $f(x)$ シグモイド関数
- W_j j 番目のニューロンからの結合値

である。

2.2 N 個のニューロンによるモデル

上述による単一ニューロンにより、リカレントニューラルネットワークを構成する。リカレントニューラルネットワークとは、各ニューロンが他のニューロンとの結合を持つばかりでなくフィードバック結合をも持つようなニューラルネットワークのことをいう。

一般に、N 個のニューロンによるモデルは以下のように記述される。

$$\tau_i \frac{du_i}{dt} = -u_i + f\left(\sum_{j=1}^n W_{ij} u_j\right) \quad (2)$$

ここに、

- $u_i(t)$ 時刻 t における i 番目のニューロンの出力
- τ_i i 番目のニューロンの時定数
- W_{ij} j 番目から i 番目のニューロンへの結合値

である。

3 個のニューロンにより構成される典型的なリカレントニューラルネットワークを図 1 として示す。

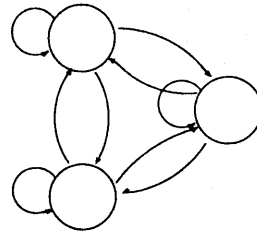


図 1: N 個のニューロンモデル (3 ニューロンの例)

3 数個のニューロンによる RNNS

RNNS の例として、まずは、数個のニューロンにより構成される“1 ペアモデル”、“2 ペアモデル”の 2 つの例を示す。

ニューロンの個数が比較的少数の場合には、演算量が少なくすむために実時間で楽音合成の可能性も高い。しかし、この場合に心配されるのは「楽音合成として実用的かどうか」ということであるが、たとえこの場合であっても楽音として十分に豊かな波形を出力することが可能であることを以下に示す。

3.1 1 ペアモデル

1 ペアモデルは、2 つのニューロンが一組のペア (以下、ペアニューロン) になっているリカレントニューラルネットワークと、もう 1 つ別の出力ニューロンとから構成されるモデルである (図 2)。

ペアニューロン内の 2 つのニューロンは、お互いに相手に結合していると共に自分へのフィードバック結合をも持つ。出力ニューロンは、ペアニューロンからの 2 つの出力を入力として受け取り、出力を

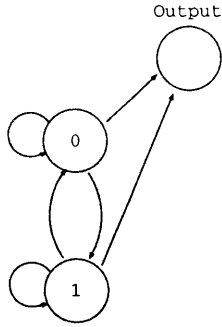


図 2: 1 ペアモデル

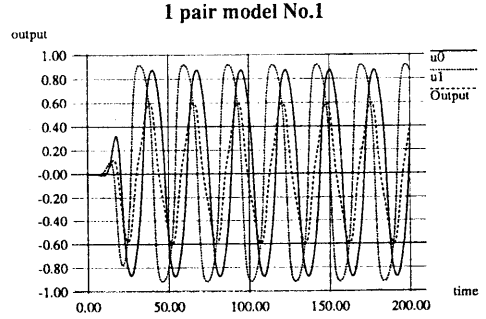


図 3: 1 ペアモデルの出力波形例 1

演算する。演算された出力が直接、音源の出力波形となる。

ペアニューロンのふるまいは、初期値・時定数・結合値の合計 8 個のパラメータで完全に記述される。パラメータによる 8 次元空間上において出力に振動解を与える領域は少なく、知られている振動解の領域としては、ペアにおけるニューロンをお互いに結ぶ結合値が反対称の場合である [Sato, 1993]。このとき、ペアニューロンの 2 つの出力は興奮状態を互いに繰り返し、一種のオシレータとして機能する。ペアニューロンからの出力をそのまま音源の出力波形として用いることも可能であるが、より複雑で豊かな出力波形を得るために、もう 1 つ別の出力ニューロンを用いた。

1 ペアモデルの出力波形の例を図 3 として示す。図 3 において、 u_0 はニューロン 0 の出力、 u_1 はニューロン 1 の出力、Output は出力ニューロンの出力、すなわち音源の出力波形、を示す。この波形は図 4 で示されるパラメータ条件において出力される。

出力波形の周波数は主にペアニューロンの時定数 τ によって決まり、波の形は主に結合値によって決まる。

次に、図 4 の条件のときよりも、ペアニューロンのフィードバック結合の値を増加したときの出力波形の例を、図 5、図 6 として示す。

このように、単純な 1 ペアモデルでありながらも、パラメータの値をいろいろ変化させることにより、さまざまな波形を出力するオシレータを作ることができる。

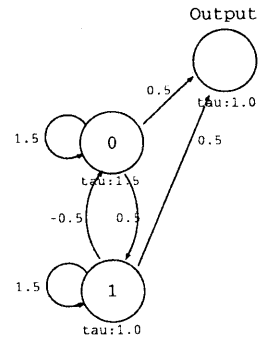


図 4: 1 ペアモデルの出力波形例 1 の各パラメータ

3.2 2 ペアモデル

2 ペアモデルは、2 つのペアニューロンと 1 つの出力ニューロンから構成される (図 7)。このモデルにおいては、オシレータが 2 つあると考えられる。各々のオシレータの周波数は各々のペアニューロンの時定数 τ に主に依存する。以上より、このモデルは FM 合成における 3 オペレータモデル (図 8) に類似していると言える。2 つのペアニューロンの周波数比率を変えることにより、さまざまな出力波形を得ることが可能となる。

2 ペアモデルの出力波形の例を図 9 として示す。この波形は図 10 で示されるパラメータ条件において出力される。

図 10 でわかるように、この例では、左右のペアニューロンの時定数の比率を 1.5 に設定している。このようにして倍音を制御することが可能となる。また、右のペアニューロンにおいて、上のニューロ

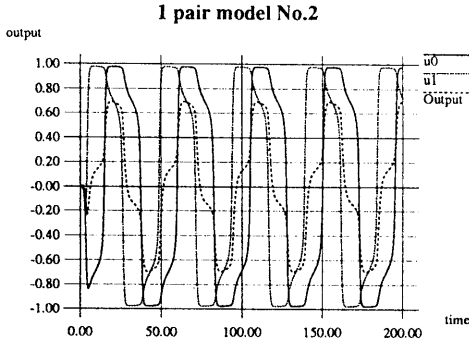


図 5: 1 ペアモデルの出力波形例 2

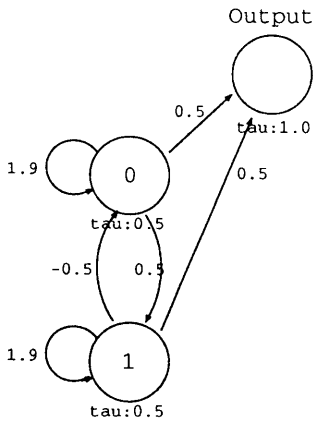


図 6: 1 ペアモデルの出力波形例 2 の各パラメータ

ンの時定数を下のニューロンの時定数に対してやや大きくしてある。これによって、右のペアニューロンの出力をわずかにゆらがせることが可能となる。

図 9 で示される音は擦弦楽器に類似しており、現存する楽器ではチェロに非常に近い。

4 学習による RNNs

前節において、ニューロンの個数が比較的少数であっても、リカレントニューラルネットワークによって現実的な楽音合成が可能であることを示した。

本節では、ニューロンの個数が十分に多い場合のモデルにおける、学習による RNNs を提示する。

ある種の構造のリカレントニューラルネットワーク

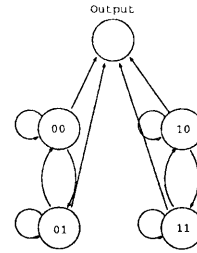


図 7: 2 ペアモデル

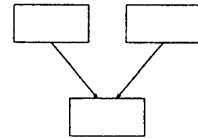


図 8: FM 合成における 3 オペレータモデル

は学習則を持ち [Pearlmuter, 1989] [Sato, 1990a]、ニューロンの個数が十分であれば与えられた波形のダイナミクスを学習することが可能である。このときの学習アルゴリズムは、よく知られているようなバックプロパゲーションにおける与えられた波形データの二次元的な丸暗記ではなく、波形の局所的な変化を徐々に学習し最終的に波形の背後にあるダイナミクスの学習に至る、というものである。

この学習則の性質は、物理的な打弦により発音されるピアノの音などに適していると考えられる。本節ではピアノの音の学習による RNNs を示す。

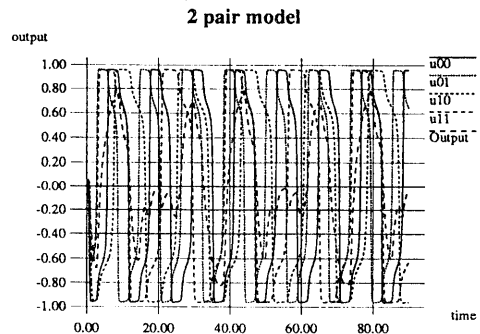


図 9: 2 ペアモデルの出力波形例

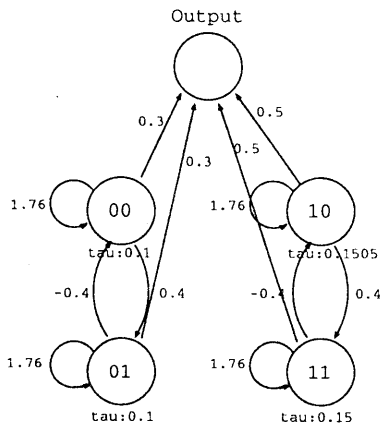


図 10: 2 ペアモデルの出力波形例のパラメータ

4.1 リカレントニューラルネットワークの構成

APOLONN[Sato, 1990c] は佐藤により提唱されたリカレントニューラルネットワークのアーキテクチャであり、多数のペアニューロンから構成されている(図 11)。

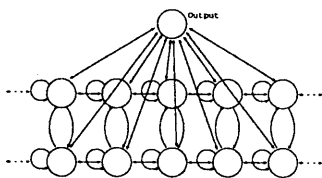


図 11: APOLONN

APOLONN の特徴をまとめると以下のようになる。

1. 結合数が比較的小さいため、学習則が比較的に簡単であること
2. 隣接するペアニューロン同士が結合していること

第 2 の特徴により、システムは非常に複雑な非線型パターンを出力することが可能となる。

APOLONN は非線型パターンに強く、また、学習則も知られているため [Sato, 1990c]、本節ではリカレントニューラルネットワークの構成として APOLONN を用いることにした。

4.2 学習データ

自然楽器音の学習データとして生ピアノを用い、A4 (440Hz) の音を 16bit, 48kHz でサンプリングした。サンプリングした全データのうち、APOLONN への学習データとしては 5016 サンプルを用いた。これは、46 周期分の波形に相当する。図 12 はサンプリングされた全データであり、そのうち、実際に学習に用いられたデータは図上で矢印にはさまれた区間である。この部分を拡大して図 13 として示す。全データのうち、比較的音量が平坦な部分を用いた。

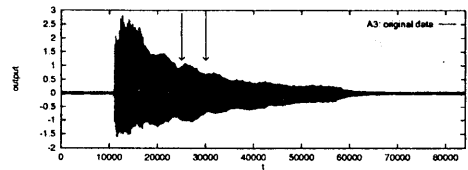


図 12: サンプリングされた全データ (A4, 440Hz)

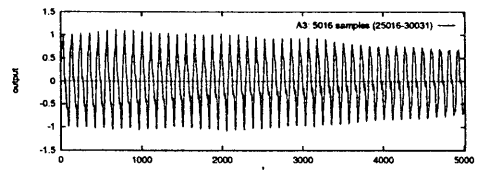


図 13: 学習データ (A4, 440Hz)

学習に用いられたデータを 3 次元位相空間に投影したものが図 14 である。遅れ時間は周期の約 1/3 とした。これにより、データが単純なものではなく、ゆらぎなどの度合いが非常に高い複雑なものであることがわかる。

4.3 学習条件

学習は、ワークステーション (Sun SPARCstation 10) 上でソフトウェアにより実行された。

APOLONN の構成としては、ペアニューロンの数を 20 個とし、各ペアニューロンの時定数は隣接するもの同士の比率を 0.9 と設定した。

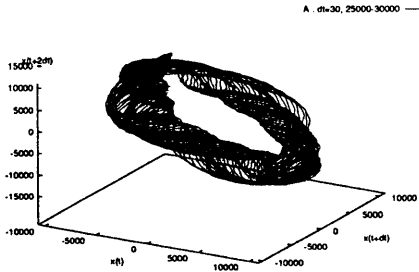


図 14: 学習データの位相図

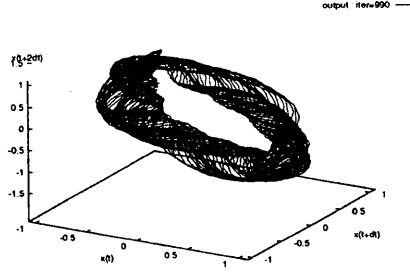


図 16: 学習後の出力の位相図

4.4 学習結果

990 回の学習の後に誤差が十分に小さくなったので学習を停止させた。図 15 に学習後の出力を示し、図 16 にその位相図を示す。

これにより、システムが複雑な時系列データを学習したことがわかる。

システムは微分方程式であるため、与えられたデータにより示された時間以降においても波形を出力することが可能である。このときの波形は学習の状態によってさまざまに変化する。これについては稿を改めて詳述することにする。

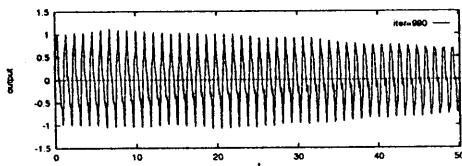


図 15: 学習後の出力

5 まとめ

リカレントニューラルネットワークによる新たな楽音合成を提示した。ニューロンの個数が比較的少数であっても現実的に楽音合成が可能であることを示した。また、ニューロンの個数が多い場合には学習により与えられた波形を再合成することが可能であり、非線型な楽音合成に新たな可能性を開いた。

参考文献

- [Murakami, 1991] Murakami, Y., and Sato, M., (1991). "A recurrent network which learns chaotic dynamics". *Proc. of ACNN'91*, pp.1-4.
- [Ohya, 1994a] 大矢健一. (1994). "ニューラルネットのダイナミクスによるリズム認知モデル". *IPJS SIG Notes*, 94-MUS-7, pp.1-6.
- [Ohya, 1994b] Ohya, Ken'ichi. (1994). "A Rhythm Perception Model by Neural Rhythm Generators". *Proceedings of the 1994 International Computer Music Conference*, pp.129-130.
- [Ohya, 1995] Ohya, Ken'ichi. (1995). "A Sound Synthesis by Recurrent Neural Network". *Proceedings of the 1995 International Computer Music Conference*, pp.420-423.
- [Pearlmutter, 1989] Pearlmutter, B. A. (1989). "Learning state space trajectories in the recurrent neural network". *Neural Computation*, 1 (2), pp.263-269.
- [Sato, 1990a] Sato, M. (1990). "A learning algorithm to teach spatiotemporal patterns to recurrent neural networks". *Biological Cybernetics*, 62, pp.259-263.
- [Sato, 1990b] Sato, M., Murakami, Y., & Joe K. (1990). "Chaotic dynamics by recurrent neural networks". *Proceedings of the International Conference on Fuzzy Logic and Neural Networks*, pp.601-604.
- [Sato, 1990c] Sato, M., Joe, K., & Hirahata T. (1990). "APOLONN brings us to the real world: Learning nonlinear dynamics and fluctuations in nature". *Proceedings of the International Joint Conference on Neural Networks*, San Diego, I, pp.581-587.
- [Sato, 1993] 佐藤雅昭. (1993). "リカレントニューラルネットワークにおけるカオスダイナミクスの学習". 「ニューラルシステムにおけるカオス」(合原一幸 編著), 第7章 pp.260-285, 東京電機大学出版局.