

# 遅延素子の層結合によるネットワークモデル A network model using delay units

城 真範<sup>†</sup>  
Masanori Shiro

<sup>†</sup> 東大生産研・産総研  
IIS, Univ. of Tokyo, National Institute of Advanced Industrial Science and Technology  
shiro@ni.aist.go.jp

## Abstract

We suggest a network consists of delay units that to average signals from pre-units, and to add a specific delay to output timeseries. By optimizing the delays in each units, the network will make a low variance timeseries from input one. We made single, two and three layers networks and tested by 4 types of inputs included chaos. We found that a network constructed by such linear units is able to cancel nonlinear waves. We would like to contribute to make a model for audio cortex in future.

**Keywords** — network, phase, delay, time series

## 1. 目的

本研究は二つの目的がある。一つは、聴覚系において信号を補償するメカニズムを提案することであり、もう一つはシグモイド関数を使わないUnitで構成されたネットワークを提案することである。以下にそれを記述する。

### 1.1 聴覚系の信号補償

人は一般に定常的な音声を聞き続けると、だんだんとそれが意識にのぼらなくなり、感じにくくなる。この現象について考えてみたい。実際の脳では意識のメカニズムが関係するため複雑であろうが、ナイーブに考える場合、これは入力された音声信号がいくつかの神経細胞を経由するうちに、だんだんとキャンセルされ、無振動の波形になることを意味するだろう。

これを模したネットワーク構造で最も簡単に思いつくものは、単純に信号の強度を減衰させるUnitで構成されたネットワークである。これはいかなる波形でも無振動の波形に変換するが、重要な問題がある。それは、非定常の波形が入力された場合は、我々はそれを鋭く感じるという事実である。すなわち非定常の波形については減衰しないモデルが必要である。そのためには任意の定常的な波形について、それをネットワークが補償的、すなわち入力波形をキャンセルするような信号変

換を学習し、出力を平滑化する枠組みが必要であろう。

実際の聴覚神経についての知見は様々に報告されているが、いったんこれを脇に置いて、ここでは入力された音声時系列を信号の強弱を変更することなしにキャンセルするようなネットワークが構成しうるかどうかについて検討したい。このためにネットワークの各Unitに課す制約は、各Unitが単に信号を遅延させる以上の機能を持たない点である。遅延量を学習によって調節することで、複雑であっても定常な入力信号をキャンセルするネットワークを構成できるならば、それは信号の定常性が崩れたときに、それを検出する回路としても機能するだろう。ただし学習アルゴリズムは本来は実際の神経回路的知見に合致したものであるべきだが、今回は神経回路との整合を必ずしも持たない工学的な最適化を使い、これについては今後の課題とする。

### 1.2 非シグモイド関数Unitの採用

上記のような目的の回路モデルとして通常思い浮かぶものは誤差逆伝搬によって学習する二層のニューラルネットワークである。ニューラルネットワークの各Unitは入力信号を加算し、シグモイド関数を通すことで信号を $[0,1]$ に規格化して出力する。シグモイド関数を通すことでUnitは非線形的性質をもち、様々な非線形的時系列に対しても学習が可能になる。ニューラルネットワークは広く使われており、実用上も十分な性能があるが、一つの問題は、学習によって最適化されたネットワークが極めて数学的に表現しにくい点である。しかも問題の大部分はUnitがシグモイド関数を使っている点そのものにあり、根本的である。

そこで今回の報告では、シグモイド関数を使わず、単に入力信号を加算平均するだけの機能をもつUnitでネットワークを構成する。これは必然的にUnitが線形的性質を持つことを意味する。本稿でいうUnitは、必ずしも現実のニューロンを意味しない。ニューロンには入出力間に非線形性があ

るが、本稿で今回報告するUnitは、加算して遅延するだけであるのでARモデルの枠を出ない。実在の脳神経との対応では、機能を持ったニューロン集団などを想定する。

## 2. 方法・条件

実験条件をそろえるため、以下では[0,1]に規格化された各10000点の時系列データを利用し、最適化手法は勾配を使わない局所最適化法であるConstrained optimization by linear approximation (COBYLA)法[1]に統一した。各Unitがとりうる遅延は0点から1000点まで(入力時系列の10%)の範囲に限定した。

サンプル信号時系列は、以下に示す4つを利用した。

- 一様乱数：メルセンヌツイスタを利用。
- 正弦波：周期  $T=628$  点。
- Logistic写像： $x_{n+1} = 3.8x_n(1 - x_n)$ で  $x_0 = 0.1$  から生成したカオス時系列。
- ヴァイオリンの音：RWC音楽データベース（楽器音）[2, 3]を利用。44.1kHzのラ音。

各Unitは、複数の親Unitからの信号を平均し、Unit固有の遅延をさせて子Unitに送り出す。この遅延量をCOBYLA最適化によって調整する。Unit数は8, 14, 20, 26, 32, 38, 44, 50とし、ネットワーク構成は1層、2層、3層とした。例えば、2層でUnit数が8であれば、図1に示すように入力Unit1つ、出力Unit1つで、各3個のUnitで構成される2層となる。図1の下側が親Unitとなり、上側が子Unitとなる。層と層のあいだは全結合される。なお、層構造をもたない全結合は、入力から出力までのあいだにループが発生し、遅延量が一意とならなくなるため、今回はネットワーク構成として採用しなかった。また、ランダム結合の場合も、入力信号を持たないUnitが発生する可能性があるため、これも採用しなかった。

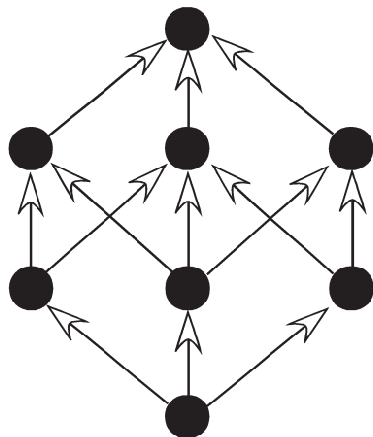


図1：2層Unit数8の場合のネットワーク構成  
実装はC++言語にて行い、最適化にnloptライブラリ

[4]を利用した。Unit数の大小によって計算時間は異なるが、最適化に要する時間は数秒から十数秒であった。

## 3. 結果と考察

以下では、出力信号の振動の度合い（平滑化の程度）をすべて分散で評価している。ただし、分散を利用する場合、下記二点の注意点がある。

1. 入力に与える各サンプル時系列は[0,1]に規格化されているが、その分散は互いに異なっている。このため、今回は入力と出力の分散の比（以下、分散比）によって、最適化の達成度合いを評価した。
2. 各Unitを経由するごとに信号が平均化処理されるため、中心極限定理の適用を受け、分散は変化する。さらに相互に結合した複数層を経由してくる信号の分散は、解析的に評価できない。このため本稿では、乱数時系列の分散の基準として示すことにした。

まず乱数列について、図2にUnit数と分散比の関係を示す。乱数列であっても、平均化処理をすることによって値の一様分布性が失われ、一定値に収束してゆく。このため、層を構成するUnitの数が増えるに連れて分散比は小さくなってゆく。合計として同じUnit数であっても、層が増えるほうが分散比が小さくなる傾向が分かる。その最低値は概ね $10^{-3}$ 程度である。

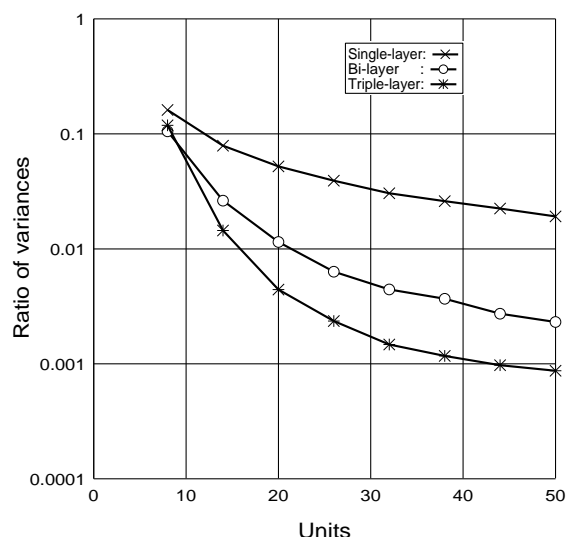


図2：ランダム時系列の最適化

次に最適化が正しく働いていることを確認するため、正弦波を入力とした。正弦波は逆位相にして元の時系列に加算するだけで信号をキャンセルできるため、原理的には4つ以上のUnitがあれば実現できる。図3はその結果である。分散比が横軸のUnit数に関係なく極めて小さな値で無秩序に揺らいでいるのは、最適化が十分に完了している

ことを示しており、また利用した最適化の精度が分散比で $10^{-6}$ 程度であることも読み取れる。本稿の実験条件で実際の聴覚との対応を述べるのは適切ではないが、この結果は、正弦波のような単純な音が容易に「気にならなくなる」という直感に矛盾しない。

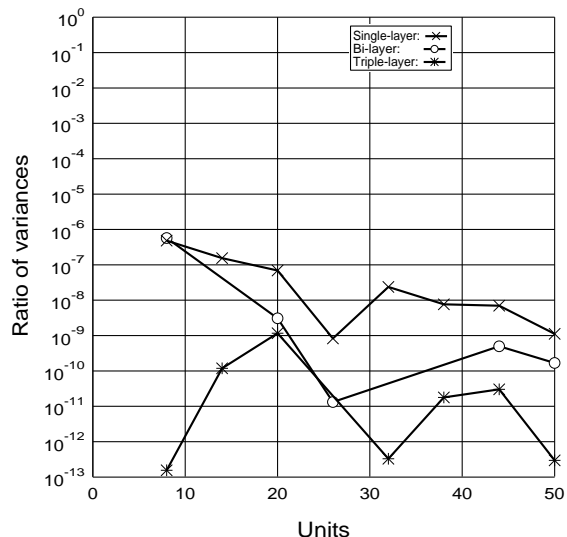


図3：正弦波時系列の最適化

次の図4は、低次元カオスとして典型的なLogistic写像を与えた結果である。図のスケールは乱数の場合と同じである。傾向は乱数の場合と同じで、多層になるほうが分散比が小さくなるが、その最低値は乱数時系列の場合よりも1/3程度小さくなっている。Logistic写像をそのまま音声に直して聞くと雑音にしか聞こえないが、提案Unitによるネットワークでは違いが検出された。

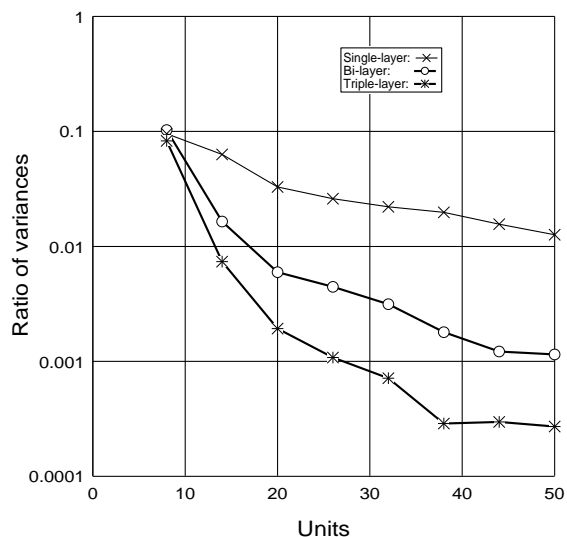


図4：Logistic写像時系列の最適化

図5はヴァイオリンの音の時系列である。これは多数の倍音を含み、かつ擬周期的であり、完全に同じ波形を繰り返さない。分散比は乱数列の場合よりも全体的に一桁程度低くなっている。これは波形の周期的な性質が効いているものと思われる。ヴァイオリンの音はこれら5つの波形の中では最も「良い音」として認知されるだろう。しかし今回の試験ではまだサンプルが少なく、その違いがどこに現れるのかは明かではない。

合よりも全体的に一桁程度低くなっている。これは波形の周期的な性質が効いているものと思われる。ヴァイオリンの音はこれら5つの波形の中では最も「良い音」として認知されるだろう。しかし今回の試験ではまだサンプルが少なく、その違いがどこに現れるのかは明かではない。

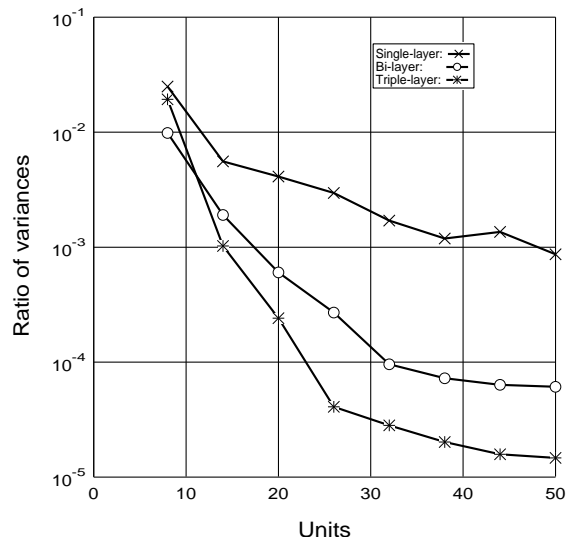


図5：ヴァイオリンの音時系列の最適化

#### 4. 今後の展開

現時点で次の4点については、既知の課題である。

1. 今回のUnitを使う限り、線形最適化の枠組みを出ない。
2. 実際の聴覚系は感覚器の時点で周波数領域への変換が行われているので、今回の研究が直接聴覚系について示唆を与えるものではない。
3. 最適化方法が非生体的である。
4. 閉路を含むネットワークに対応できていない。

特に、1.について詳述する。

人の声やいくつかの楽器の音が10次元程度の決定論的カオスであることが指摘されている。例えばクラリネットの音や、本稿で使用したヴァイオリンの音もそうである。一方、単なる風の音や波の音や電話の呼び出し音は決定論的カオスではない、もしくは非常に高次元のカオスであるため統計的手法では検出できない。楽器の音は人間が歴史をかけて良い音を追求してきた結果であり、良い音と低次元のカオス性が何らかの関係を持っている可能性がある。しかしながら今回実験に使ったLogistic写像などはカオスではあるが音にして聞いても雑音にしか聞こえない。カオスの中でも擬周期的であり、特別の特徴を持つ時系列が「良い音」の評価に関係していると思われるが、現状の単なる線形的な加算モデルでは、こうした視点からの研究は原理的に難しい。

しかし、提案手法は、Unitの性質を自由に変えられるところに発展性がある。現在は（最適化の変数を抑える意味もあって）ただ親Unitからの信号を加算平均して遅延するだけであるが、親Unitからの信号を各係数に割り当てて多項式演算をして出力するなどといった改良をすることで、上記1.は克服できる可能性がある。Unitをこのように多項式演算に対応させることは、一見認知的なメカニズムの解明からは遠ざかるかもしれないが、数理的な扱いを容易にするため、詳細に検討することで認知的なメカニズムの解明とモデル化へ貢献できる可能性もある。

直接に実際の聴覚系に近づける方法としては、信号をそのままではなくフーリエ変換してから入力にし、位相と強度の係数行列を最適化してゆく等の改良が必要である。また、閉路に対応する必要もある。入力を擬周期波形に限って、比較のため波形の周期を統一する必要もあるかもしれない。

## 参考文献

- [1] Powell, M. J. (1994). A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in optimization and numerical analysis* (pp. 51-67). Springer Netherlands.
- [2] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, *Proc. 4th Int. Conf. Music Information Retrieval (ISMIR 2003)*, 229 (2003).
- [3] M. Goto, *Proc. 18th Int. Congress on Acoustics (ICA 2004)*, 553 (2004).
- [4] <http://ab-initio.mit.edu/wiki/index.php/NLopt>