# ニューラルネットワークを用いた自動採譜への試み

2N−4

秋田 真彦　増山 繁
豊橋技術科学大学知識情報工学系

## 1 はじめに

本研究は4層ニューラルネット [5] を用いて2種類の楽器の単音の混成音から各楽器の単音を識別、分離し、音高を同定することを目的とする。[1] では、ニューラルネットを用いた楽音識別の研究がなされている。ところが、音高の識別まで試みようとすると、そのままではネットワークの規模が大きくなり、多くの学習時間を要し、また、汎化できなくなる可能性がある。そこで、本研究ではニューラルネットの規模を縮小するため入力に対して主成分分析 [2, 6] を用いるアプローチをとり、以下の手続きでクラスター分割を容易にさせると同時に、入力情報を圧縮する。

1. テンプレートのための音声データをメモリ上で合成し、一つの音声データに対し、10個の任意に定めた解析点から256点フーリエ変換する。
2. 一つの音声データを1行256列の行列に対応させると、(10× 音声データ数)行 256列の行列ができる。これに対し主成分分析を行ない、256行256列の固有ベクトル $Q$ を得る。また、$Q$ を寄与率の高い順に並び変えた ベクトル $Q'$ を作っておく。
3. 一定の累積寄与率を定めておき、$Q'$ のうちその累積寄与率を実現するまでに影響する $Q'$ の行数 $n$ だけコピーした $n$ 行256列の行列 $Q''$ を作る。
4. $Q''$ に 1行256列の音声データの転置 (256行1列の行列) を掛けると、1行 $n$ 列の行列を得る。これは音声データ群を主成分分析し、(1- 累積寄与率) ×100% の誤差範囲内におさまる第 $n$ 主成分までの座標である。
5. こうして得られた $n$ 次元の座標を入力とし (入力素子は $n$ 素子)、ニューラルネットワークを組む。
   この結果、例えば2章に示すように、通常なら256入力のところを2入力に削減することができた。

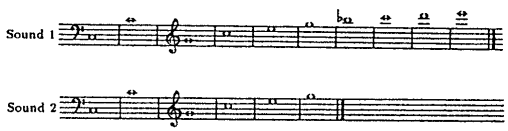
本稿では、第2章で単一楽器の単音から音高を識別する実験を示し、3章で2つの楽器の単音の混成音から楽器を分離、音高を識別する実験を示す。

なお、類似した実験例では、阿部 [1] らがニューラルネットを用いて単音からの楽器識別を行ない、片寄 [3] はアコースティックアルゴリズムコンバイラを用いて音源分離および採譜を行なっている。また、田村ら [4] は4層ニューラルネットを用いて雑音抑制の解析を行なっている。

本研究ではまず、実際の楽器音を用いず、メモリ上で合成した波形を用いて実行可能性を検討した。メモリ上での合成、各係数での解析、ニューラルネットワークのプログラムはすべてSun SPARC Station 1上でC言語を用いて作成した。

## 2 単一楽器の単音から音高を識別する実験

本章では単一楽器の単音から音高を識別する実験を示す。




Sound 1 : $S_1(t) = \sin(ft) + \sin(2ft) + \sin(3ft)$
Sound 2 : $S_2(t) = \sin(ft) + \sin(3ft) + \sin(5ft)$
ただし、$f$ : 基本周波数、t: 時間

図1 : 音声データ
音声データの組合せ数は $10 + 6 = 60$ 個

図1のようなシンセサイザーで合成した2種類の楽器の単音の混成音データを想定し、メモリ上で合成した。一つの音声データに対し、それぞれ10個の任意に定めた解析点から、周波数スペクトル解析を行なう。実験機器に33KHzサンプリング可能なものがあるので、それに合わせ、周波数スペクトル解析は256点FFTとした。さて、以上で得られた係数を1章で述べた方法で2次元に射影する。この結果、クラスターがまとまっていることから、ニューラルネットによる識別が容易であろうと思われるので、以下ではこれを採用した。

また、それぞれのカテゴリが重なっていないので、2入力で十分であることを示している。

Extracting a Score from Polytone using Four-layerd Feedforward Neural Networks
Masahiko AKITA,Shigeru MASUYAMA
Toyohashi Univ. of Tech.

## 2.1 学習結果

ここでは、学習を容易にするために、各入力素子への入力を便宜上700倍し、実数値入力を0,1入力に変換した。すなわち、入力層1400素子、出力層16素子、4層で中間層の素子数を20素子とするネットワークを組み、学習させる。学習結果として図2に、横軸に学習回数、縦軸に誤差をとったものを示す。

なお、このように入力素子を増加させても、実数値を入力する入力層2素子のネットワークよりも学習が早いという結果を得ている。また、0への閾値を$10^{-4}$として、誤差が0となった学習回数は2200回で、時間にして約1日である。

## 2.2 汎化の可能性

上のネットワークに対して新たな未学習パターンを数十個用意し、学習させた場合の、未学習パターンに対する誤差を図3に示す。ここで、学習時間に2週間ほど費やしているが、この結果、汎化の可能性が高いと思われる。

## 3 2つの楽器の単音の混成音からの楽器分離、音高識別の可能性

図1の混成音の場合、音声データの組合せ数は10×6＝60個になるが、この周波数スペクトル解析に主成分分析を適用する方法によると、256次元の入力が寄与率0.8で8次元にまで削減できている。

そこで、2章と同様に各入力素子への入力を便宜上700倍し、実数値入力を0,1入力に変換し、学習させる。すなわち、入力層700×8＝5600素子、出力層16素子、4層で中間層の素子数を20素子のネットワークを組み、学習させる。学習結果として図4に、横軸に学習回数、縦軸に誤差をとったものを示す。2章と同様に汎化をさせれば、楽器、音高の識別が可能であると思われる。

## 4 結論・今後の課題

本稿では、ニューラルネットワークを用いて周波数スペクトルを手がかりにして楽器を識別、音高を同定する際に、周波数スペクトル解析に主成分分析を適用する方法により、ネットワークの規模を削減できることを明らかにし、単一楽器の単音から楽器および音高を識別できることを明らかにした。

また、2つの楽器の単音の混成音から楽器および音高を識別できる可能性が高いことを示した。今後の課題として、今の方法のままでは次元数が大きくなると入力素子数が多くなり学習時間が長くなるので、これを解決するために効率の良い入力の与え方、学習方法を考案する必要がある。また、将来的な課題として、この方法のままだと混成音を構成する音数が増加したり、楽器数が増加すると学習パターン数に組合せ的爆発が起こるので、その問題も解決していかねばならないだろう。

## 謝辞

本研究を進めるにあたって有益な助言を賜わった、豊橋技術科学大学情報工学系の舟橋賢一 先生に深謝の意を表する。

## 参考文献

[1] 阿部 素詞, 阪口 豊, 中野 馨,“楽器音の学習認識システム”, SICE 学術講習会 91 予稿集, p.605-606 (1991)

[2] 奥村 晴彦,“C言語による最新アルゴリズム事典”, 技術評論社, p111-113(1991.9)

[3] 片寄 晴弘,“音楽感性情報処理に関する研究”, 大阪大学基礎工学部博士論文 (1991.1)

[4] Shin'ichi Tamura,Alex Wabel, “Noise reduction using connectionist models”, IEEE ICASSP, p.553-556(1988)

[5] 中野 馨, 飯沼 一元, ニューロンネットグループ, 桐谷 滋, “入門と実習ニューロコンピュータ”, 技術評論社,(1991.1)

[6] 柳井 晴夫, 高根 芳雄,“多変量解析法”, 朝倉書店 (1979.1) 第3刷

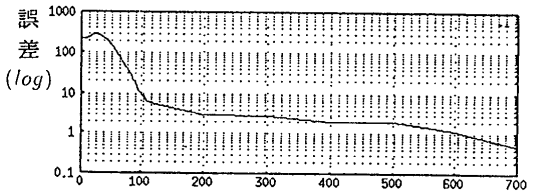


図 2: 単音の学習曲線　　学習回数 (回)

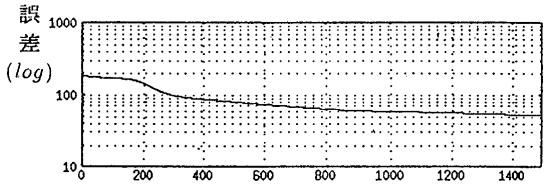


図 3: 未学習パターンに対する誤差　　学習数 (パターン)

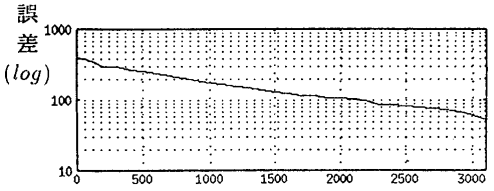


図 4: 単音の混成音の学習曲線　　学習回数 (回)