

# ニューラルネットワークを用いたジャズセッションシステム 4R-5 —ニューロミュージシャン—

渡辺和之 西嶋正子 柿本正憲 村上公一  
(株)富士通研究所

## 1. はじめに

我々は、人間の感性のような理屈では言い表わしにくい情報を扱うことができるコンピュータの開発を目指している。音楽は人間の感性表現の中の一つであり、工学的にも興味深い分野である。音楽は長い歴史の中で構築されてきた理論的な面をもっているため、人工知能的なアプローチからの研究が盛んに行われているが、感性的な情報をどのように処理するかという問題を抱えている。

ニューラルネットワークは人間の直感的な処理を模倣できるのではないかと考えられている。特に人間の直感的な処理を必要とする音楽の形態に「セッション」と呼ばれる即興の掛け合い演奏があり、演奏者は互いに影響を受け合いながら演奏を展開してゆく。そこで我々は、セッションにおいてニューラルネットワークを感性的な情報の処理を行う部分に適用し、人間対人間のセッションを、相手をコンピュータに置き換えることによって実現させることを試みた。

まず、我々はリズムに関する情報を扱うことにより、人間とコンピュータが実時間でドラム間のセッションを行うことができるシステム「ニューロ・ドラマー」を開発した[1]。次に、メロディに関する情報を扱うことにより、メロディ楽器間のセッションを行うことができるシステム「ニューロ・ミュージシャン」を構築した。

本稿では、後者の「ニューロ・ミュージシャン」についてその概要および構成について報告する。

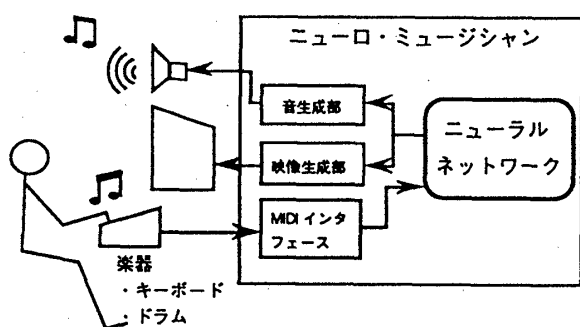


図-1 システム構成図

## 2. システムの概要

32ビットのパーソナルコンピュータ (FMR-70またはFMTOWNS) 上に構築したシステム「ニューロ・ミュージシャン」の構成図を図-1に示す。なお、パーソナルコンピュータと楽器間の通信にはMIDI (Musical Instruments Digital Interface) を用いた。

### 2.1 モーツァルトソナタシステム

過去の音楽作品のメロディをニューラルネットワークに学習させることによって、新しいメロディを生成させる実験を行った。今回は音楽作品数曲について実験を行ったが、モーツァルトのピアノソナタ第11番を中心に行ったので、このシステムをモーツァルトソナタシステムと呼ぶことにする。学習と生成の過程を図-2に示す。まず学習の過程では、入力メロディに対する期待する出力メロディのペアを数種類ネットワークに学習させる。そして生成の過程では、学習済みのネットワークを使用する。学習したメロディを入力したときには学習した通りの出力メロディを返してくるが、学習していないメロディを入力したときには新しいメロディを返してくる。その新しいメロディが学習したメロディの傾向を反映していれば、このネットワークはメロディの特徴をうまく抽出することができたと言うことができる。このシステムで使用したネットワークは3層の階層型ネットワークであり、ノード数は入力層24、中間層30、出力層24である。また、扱った情報は1小節のメロディの音の高さ、音の強さ、音の長さの3種類である。

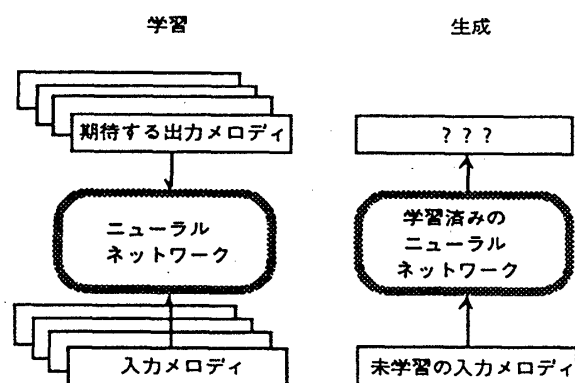


図-2 学習過程と生成過程

A Jazz Session System based on Neural Networks - Neuro-Musician -

Kazuyuki WATANABE, Masako NISHIJIMA, Masanori KAKIMOTO, Koichi MURAKAMI  
FUJITSU LABORATORIES LTD.

2. 2 ジャズセッションシステム

ジャズにおけるセッションの場合は、8小節のメロディの掛け合い演奏を実現した。従って上述のモーツァルトソナタシステムで扱った音の高さ、音の強さ、音の長さの3つの情報以外に、8小節のメロディ全体の輪郭、音の分布情報、音を出すタイミングの更に3つの情報を扱った。

(1) メロディの輪郭

我々はメロディ全体の輪郭を扱うために、「ゲート」というものを定義した。ゲートとは4分の4拍子の場合、最初の音、3拍目の音、最後の音と1小節につき3つの音の高さの情報である(図-3)。8小節分の入力ゲート情報と期待する8小節分の出力ゲート情報のペアを1段目のネットワーク「ゲート生成用ネットワーク」に学習させる。このゲート生成用のネットワークは3層の階層型ネットワークであり、ノード数は入力層48、中間層60、出力層48である。

(2) 音の分布情報

音の分布情報を扱うために、演奏された8小節のメロディの中の音の数、およびそれらがどの音域に分布しているかという情報を抽出する。そして、その抽出された分布情報とゲート情報を合わせたものを入力情報とし、期待するメロディを出力情報として2段目のネットワーク「メロディ生成用のネットワーク」に学習させる。このメロディ生成用のネットワークは2小節ごとに処理を行うため4個のネットワークを用いており、各ネットワークは3層の階層型ネットワークで、ノード数は入力層24、中間層40、出力層32である。



▲：抽出するゲート

図-3 演奏データにおけるゲート

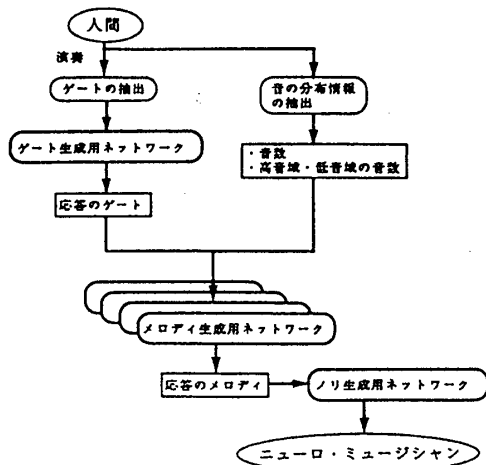


図-4 ジャズセッションシステム

(3) 音を出すタイミング

実際のジャズの演奏は、メトロノームのような正確なリズムの中で行われるわけではなく、音を出すタイミングを微妙にずらしながら行われる。このタイミングのずらし方を「ノリ」と呼ぶことにする。ノリは演奏されるリズムの形にかなり依存しており、また演奏者によっても違いが見られる。従ってリズムの形とノリの関係を測定を行い、その情報を3段目のネットワーク「ノリ生成用ネットワーク」に学習させる。このノリ生成用のネットワークは3層の階層型ネットワークであり、ノード数は入力層8、中間層8、出力層13である。

(1) ~ (3) の過程を示したジャズセッションシステムの概略図を図-4に示す。

3. 実験

「ニューロ・ミュージシャン」として2種類のシステムを用いて実験を行った。まずモーツァルトソナタシステムでは、10数種類のメロディのペアをニューラルネットワークに学習させるだけで、学習したメロディの傾向を反映している新しいメロディを得ることができた。次にジャズセッションシステムでは、実際にプロのジャズピアニスト2名に学習データの作成と学習を済ませたシステムを用いたジャズセッションを依頼し、被験者になってもらい評価を行った。彼ら曰く、「ジャズの初級者程度」ということであったが、時々面白いメロディを返してくるため自分たちの演奏も良い影響を受けるそうであった。また2種類のシステムともセッションが実時間で実現できた。

4. まとめ

今回の「ニューロ・ミュージシャン」の構築により、人間とコンピュータがメロディのセッションを行うことが可能であることが確認された。ただし、今回の場合、扱ったメロディが単旋律(同時に出る音が1種類の旋律)であったり、ジャズセッションシステムではノリを考慮したものの、基本的にはリズムの細かさが16分音符単位までとしたために、生成したメロディの単調さは避けられなかった。今後は、より自然な演奏となるようにシステムの改良を行うとともに、音楽における人間の感性情報処理の研究を進め、最終的にはコンピュータが人間の良きパートナーとなるようなシステムの開発を進めていく予定である。

[参考文献]

[1] 西嶋他: "ニューロ・ドラマー", 計測と制御, Vol. 30, No.4, pp344-347 (1991)