

学習するセッションシステム： 演奏者の振る舞いのモデルの獲得

浜中 雅俊 後藤 真孝 大津 展之
筑波大学 電子技術総合研究所
{hamanaka, goto, otsu}@etl.go.jp

本稿では、実在する人間の演奏者を模倣した仮想演奏者と人間の演奏者がインタラクションできるようなジャムセッションシステムについて述べる。従来の多くのシステムでは、仮想演奏者がどのような振る舞いをするかは、アドホックなルール群等により決定されていたため、ルールのパラメータの調整により異なる個性を設定できても、実在の演奏者の振る舞いを模倣させることは困難であった。本研究では、演奏者全員の演奏履歴と、ある演奏者が現在弾いている演奏との関係を統計的に学習することにより、演奏者の振る舞いのモデル(相手の演奏に対してどのような演奏をするかを定めるモデル)を獲得する手法を提案する。ギタートリオの演奏を対象とし、上記のモデルの学習機能を備えたセッションシステムを、入出力にMIDIを用いて実装した。実験の結果、実際にMIDIギターの演奏記録から、任意の一人の振る舞いのモデルを学習可能なことを確認した。

A Learning Session System: Statistical Modeling of Player's Reactions

Masatoshi Hamanaka Masataka Goto Nobuyuki Otsu
Univ. of Tsukuba Electrotechnical Laboratory
Mbox 1704, 1-1-4 Umezono, Tukuba, Ibaraki 305-8568 Japan

This paper describes a jam session system that enables a human player to interplay with virtual players, each of which imitates musical reactions of some human player. Previous session systems have not been able to imitate such reactions, although these have parameters for altering a way of reacting. Our system can obtain the reaction model of a human player, which means a characteristic way of reacting to the other players, by learning the relationship between MIDI data which the player listens to and MIDI data improvised by the player. The experimental results show that the proposed system dealing with a guitar trio can learn the reaction model of any participant from the MIDI recording of a session.

1. はじめに

ジャムセッションでは、演奏者が「相手の演奏に対してどのような演奏で反応するか」が重要である。本研究では、それを演奏者の振る舞いと呼び、実在する人間の演奏者の振る舞いをモデル化し、それに基づいてセッションできるような仮想演奏者を実現することを目標としている。そのような振る舞いの模倣が可能となれば、親しい演奏

者や、自分よりも演奏能力の高い演奏者、既に亡くなった演奏者のモデルを用いた仮想演奏者といつでもインタラクションすることができるし、自分自身のモデルを用いた仮想演奏者とジャムセッションを行うことも可能となる。

従来のジャムセッションシステム[1][2][3]では、人間の演奏に追従した演奏を仮想演奏者にさせる

ことに主眼がおかれていたため、仮想演奏者に個性を持たせるには至らなかった。一方、文献[4][5]では、個性データベースと呼ばれるルール群を導入することにより個性の違いを設定することを可能としていた。文献[6][7]では、システム外部から変更可能なパラメータを複数用意することにより、各演奏者が主導権を握る程度を様々に変化させることを可能としていた。しかし、これら[4]-[7]は、パラメータやルールの調整を行うことにより、異なる振る舞いの仮想演奏者を設定することはできても、人間の演奏者の振る舞いを模倣するようなモデルを設定することは困難であった。

これに対し本研究では、模倣しようとしている演奏者がどんな演奏に対してどんな即興演奏を行ってきたかを、統計的に学習することにより、演奏者の振る舞いのモデルを獲得することを可能にする。つまり、セッションの演奏を記録したMIDIデータさえあれば、それに参加した任意の演奏者の振る舞いのモデルを学習できるという特長をもつ。

このような演奏者の振る舞いの模倣は、演奏者が聴取した入力演奏から演奏者が出力する演奏を求める問題として形式化できる。しかし、演奏の入出力関係をMIDIデータなどから直接、統計的に求めようとすると、膨大な量のデータセットが必要となるばかりでなく、複雑すぎてモデル化が困難となる。本研究では、このような問題を解決するため、入力と出力の間に、2つの主観空間(主観を表す空間)を導入することを考えた。演奏者の振る舞いのモデルは、入力側の主観空間から出力側の主観空間への関数をRBFネットワークを用いて学習することにより獲得できる。仮想演奏者は、得られた関数を用いて、演奏から受ける印象から演奏意図を決定し、演奏を生成する。

本研究では、文献[6]で後藤らが提唱しているセッションのモデルをさらに発展させ、人間の演奏者とその振る舞いを模倣した仮想演奏者とが置換可能となるシステムを、人間の演奏者、仮想演奏者あわせて3人がギターでセッションするという形式で実装した。3人の人間の演奏者のMIDI演奏記録をもとに学習を行ったところ、それぞれの演奏者の振る舞いのモデルが獲得できた。また、得られたモデルを用いて仮想演奏者にセッションをさせたところ、ソロや伴奏を交代しながら、モデルに基づいた振る舞いをする事が確認できた。

2. セッションシステムの概要

本セッションシステムは、仮想演奏者が人間の演奏者と対等な立場ですべての演奏者の演奏を聴き、その演奏に反応するとともに自己主張することができるよう、文献[6]の全プレイヤーが対等なセッションモデルに基づいて構成されている。本研究では、そのモデルをさらに発展させ、各演奏者が人間であるか仮想演奏者であるかが自由なだけでなく、人間の演奏者に入れ替わってその振る舞いを模倣した仮想演奏者に演奏させることが可能となるようにしている(図1)。仮想演奏者は、自分を含めたすべての演奏者の演奏を聴き、その演奏に対して自分の振る舞いを決定するため、相手が人間であるか仮想演奏者であるかは問わない。

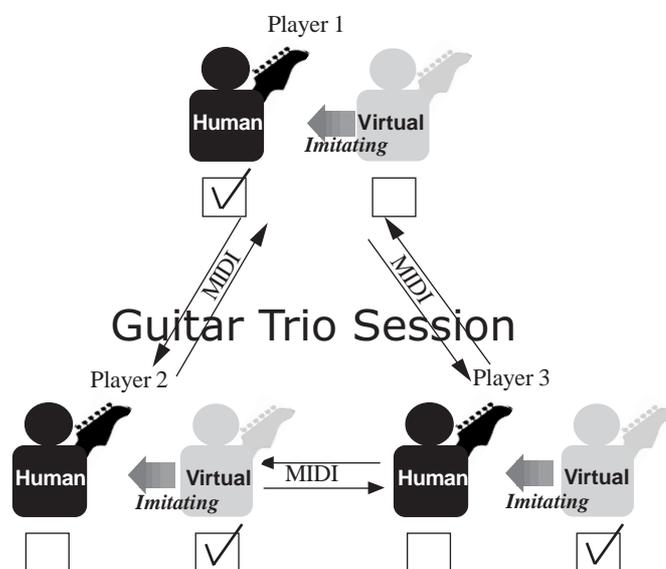


図1：人間の演奏者と仮想演奏者とが置換可能なセッションのモデル

本研究では、以上のモデルをギタートリオのセッションに適用した。すべての演奏者にギターという同一の楽器を演奏させることにより、仮想演奏者の聴取過程でも複数の相手の演奏をまったく対等にとらえることができる。3人の演奏者は12小節1コーラスの典型的なブルース進行で、ソロや伴奏を繰り返しながら演奏する。調はA、テンポは120で固定とする。

2.1 演奏者の振る舞いのモデル

本研究では演奏者の振る舞いのモデルを獲得するために「演奏印象」と「演奏意図」という2つの主観的な概念を導入した。演奏印象とは、演奏

を聴くことにより受ける印象である。たとえば、ジャムセッションでは、各演奏者の演奏を聴いて「存在感がある」とか「安定している」という印象を受ける。演奏意図とは、次にどのような演奏をしたいかである。演奏意図の変化が少なければ、似通った演奏を出力し続けることになる。

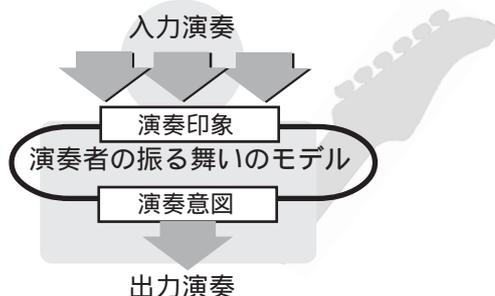


図2：演奏者の振る舞いのモデル

演奏者の振る舞いのモデルは、演奏印象を演奏意図に変換するものである（図2）。演奏者の振る舞いのモデルは、演奏印象を表す**印象空間**から演奏意図を表す**意図空間**への関数として獲得することができる。印象空間は、演奏から受ける印象を表すのに適していると思われる言葉（印象語と呼ぶ）を座標軸に持つ空間である。ある時刻において、ある1人の演奏者から受ける演奏印象は、印象空間上の1点で表される。以下、これを**印象ベクトル**と呼ぶ。印象ベクトルは、入力演奏から得られる物理的な特徴量を座標軸に持つ空間（物理特徴空間）から印象空間への写像によって決定される。このとき、演奏の物理的な特徴も物理特徴空間上の1点で表されることになり、これは**物理特徴ベクトル**と呼ぶ。一方、意図空間は、仮想演奏者が演奏を生成するとき使用する1小節から8小節の長さの演奏パターンを主観的類似度に基づき配置した空間で、ある1つの演奏パターンは意図空間上の1点で表される。以下、これを**意図ベクトル**と呼ぶ。

2.2 振る舞いのモデルの学習

演奏者の振る舞いのモデルは、人間の演奏者がセッションを行っている間に刻々と変化する印象ベクトルと意図ベクトルの組から統計的に学習することで獲得できる。印象ベクトルと意図ベクトルは、演奏者の内部状態を表すもので直接観測することはできない。そこで本研究では、振る舞いのモデルを得るために、以下のような3段階の処理を考案した(図3)。

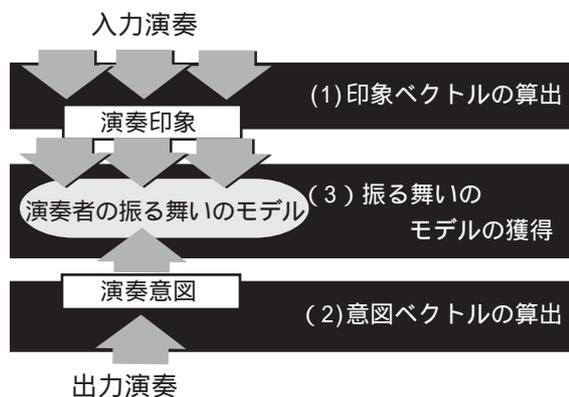


図3：振る舞いのモデルの学習のしくみ

(1)印象ベクトルの算出

演奏者が聴取した演奏のMIDIデータから物理特徴ベクトルを求め、それから印象ベクトルを算出する。このような写像を行うためには、演奏者が「ある物理的特徴をもった演奏を聴いたとき、どのような印象を受けるか」を推定する必要がある。ここでは、演奏の物理特徴ベクトルと心理実験の結果得られた印象ベクトルとの間の相関関係を正準相関分析を用いて求めることにより、印象ベクトル算出のための写像を得ている。印象ベクトルの算出は、セッションに参加しているすべての演奏者に対して別々に行うため、3つの印象ベクトルが得られる。

(2)意図ベクトルの算出

演奏者が出力した演奏のMIDIデータから物理特徴ベクトルを求め、その演奏を生成するような意図ベクトルを算出する。このような写像を行うためには、演奏者が「ある演奏を出力したとき、どのような演奏意図であったか」を推定する必要がある。しかし、この推定は、演奏者が演奏意図をもとに演奏を生成する過程の逆を実現しなければならないため、一般に難しい。

そこで本研究ではまず、出力演奏パターンの主観的類似度を心理実験で求め、得られた類似度から意図空間を構成することを試みる。このとき、ある演奏パターンは意図空間上の1点と対応づけられる。そして、各出力演奏パターンの意図ベクトルとその物理特徴ベクトルとの間の相関関係を正準相関分析を用いて求めることにより、意図ベクトル算出のための写像を得ている。

(3)振る舞いのモデルの獲得

演奏者の振る舞いのモデルを、3つの印象ベクトルの時系列から意図ベクトルへの関数として獲得する。本研究では、3つの印象ベクトルの過去12小

節の履歴を、RBFネットワークの入力ベクトル上に展開することにより、時系列を扱った学習を可能にした。

2.3 仮想演奏者の構成

仮想演奏者は、1)演奏印象算出部、2)演奏意図決定部、3)演奏生成部、の3つからなる(図4)。以下、3節、4節、5節でその詳しい内容について述べる。

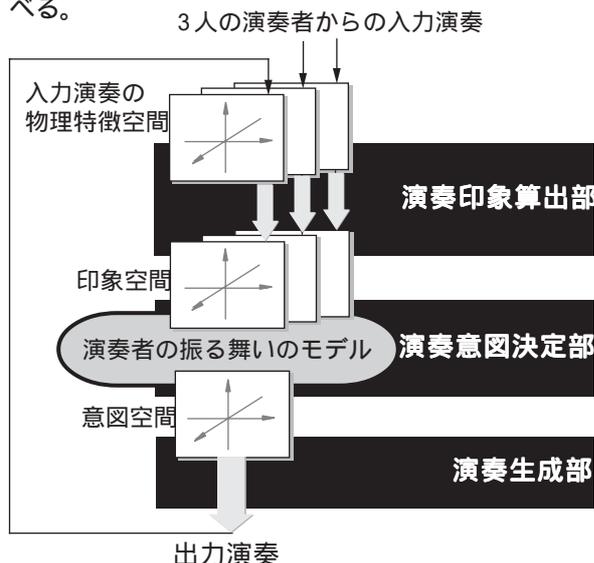


図4：仮想演奏者の構成

3. 演奏印象算出部

演奏印象算出部は、全演奏者の演奏MIDIデータから物理特徴ベクトルを求め、印象ベクトルを算出する部分である。そのとき用いる写像は、心理実験の結果得られた印象ベクトルと物理特徴ベクトルとを正準相関分析することにより求める。正準相関分析(Canonical Correlation Analysis)とは、2組の計測ベクトル間の相関関係を分析する場合に用いる手法で、複数の変数からなる2変数群(ここでは印象ベクトルと物理特徴ベクトル)それぞれについて線形合成変数を求め、2つの合成変数の相関(正準相関)がもっとも大きくなるように重みをつけるものである。このとき、2つの合成変数の相関をできる限り大きくするためには、印象ベクトルと相関の高い物理特徴量を見つける必要がある。しかし、どのような物理特徴が、演奏から受ける印象に大きく影響しているかは明らかでない。そこで本研究では、様々な種類の物理特徴量を用意することにした。

写像を求める際、印象ベクトルと物理特徴ベクトルは、できる限り低次元であることが望ましい。印象ベクトルの低次元化は、演奏者の振る舞いのモデルの学習を可能にするために必要である。ここでは、主成分分析を行い重要な印象語を選択することにより印象ベクトルを低次元化した。主成分分析(Principal Component Analysis)とは、多変量の計測値から変量間の相関を無くし、しかもより低次元の変量によって元の計測値の特性を記述するための手法である。一方、物理特徴ベクトルの低次元化は、不要なデータを取り除き純化することと、演奏印象算出部で写像を求める際の計算時間を短縮する意味をもつ。物理特徴ベクトルは、正準相関分析を繰り返し行い、正準得点の大きさによって物理特徴量の取捨選択することにより低次元化した。

3.1 印象ベクトルの獲得

(1)心理実験

被験者(音楽経験者)に演奏を聴かせながら、1/12小節ごとに、演奏を表す10語(安定感、異質感、開放感、緊張感、堅実感、重厚感、爽快感、存在感、平凡感、躍動感)の印象語がふさわしいかどうか7段階で評定させる実験を行った。用いた演奏は、ギタリストがテンポ120で行ったセッションのMIDI演奏記録(長さ12コーラス)の、ある1人分のパートの演奏である。被験者は評定結果を、実験用に作成したアプリケーションの画面上に次々と入力する。演奏を1回聴いただけでは、すべての時刻における評定を入力することは難しいため、以上の操作は繰り返し行なう。その際、前回までの評定結果は、画面上に表示される。

実験の結果、録音演奏から1728個(12コーラス×12小節×12個)の印象ベクトルを得た。被験者がセッションの展開を記憶して評定を行なうことを防ぐため、録音演奏を繰り返し聴く前に、ジャンルが異なる音楽CDを5分以上聴かせた。しかし、それでもセッションの初めや終わりは、展開を記憶してしまいやすいので、録音演奏の前後1コーラス部分を除いた、1440個(10コーラス×12小節×12個)の印象ベクトルを用いることにした。

(2)印象語の選択

印象ベクトルの主成分分析を行なったところ、累積寄与率が84%までの主成分が3つ得られ、第1、第2、第3主成分がそれぞれ、存在感、躍動感、

重厚感に近接していた。したがって、この3つの印象語でほぼ印象空間全体を表現できることがわかった。以下、この3つの印象語を座標軸に持つ空間を印象空間とし、印象ベクトルも3次元で表すことにする。

3.2 入力演奏の物理特徴ベクトルの獲得

(1) 様々な物理特徴量の抽出

独立な物理特徴量を数多く抽出するため、演奏者の印象ベクトルに影響を与えていると考えられる過去12小節の演奏を、1小節ごとに分割し、小節内のすべての音に対し、1/48小節の分解能でペロシティー、音高、ピッチベンドの値を求め、それぞれ音量フィルタ、音域フィルタ、Noteフィルタにかけた。各フィルタはそれぞれ、ペロシティーや音域、コードの主音からの音程が、ある範囲にあるときのみ1を出力することにより、様々な物理特徴の分離、抽出を可能としている。得られた値について小節内で、平均値、最大値、最小値、傾きを算出し、それらを物理特徴量とした。物理特徴量は全部で663552個生成された。

(2) 物理特徴量の取捨選択

物理特徴ベクトルと印象ベクトルとの正準相関分析を行うごとに、一番正準得点が低かった物理特徴量を1つ削除し、新たな物理特徴量を1つ加える操作を繰り返し、印象ベクトルと相関の高い物理特徴量を選択する。はじめに物理特徴ベクトルに与えられる物理特徴量は100個である。新たに追加する物理特徴量は、物理特徴ベクトルに残っている99個の物理特徴量のいずれに対しても相関が低いものとした。これは、相関の高い物理特徴量を加えても正準相関の値が変化しないだけでなく、相関行列の状態が悪くなり逆行列が求められなくなるなどの問題が起こるためである。本研究では正準相関分析を1990656回(663552×3回)行なったところ、正準相関の値が0.9を超えたため、そこで終了した。

3.3 演奏印象の獲得

印象ベクトルと物理特徴ベクトルの正準相関分析により、両者の写像が得られたため、新たな演奏が入力された場合でも、物理特徴ベクトルから印象ベクトルを求めることが可能となった。

図5は、存在感について、被験者の評定値とシステムが算出した演奏印象とを、比較した結果であ

る。92パーセントの部分で両者の差は1以内であった。システムは、1/12小節ごとに印象ベクトルを算出している。

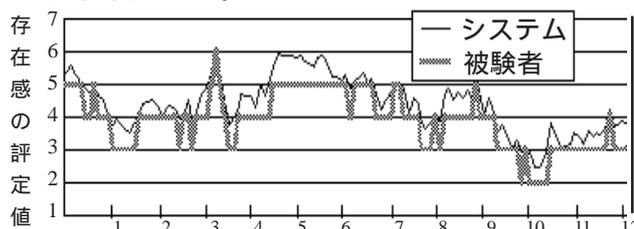


図5：被験者とシステムが示した演奏印象 [小節]

4 演奏意図決定部

演奏意図決定部は、人間の演奏者の振る舞いのモデルの学習、および得られたモデルを使って仮想演奏者の意図ベクトルを決定する部分である。演奏者の振る舞いのモデルの学習はRBFネットワークを用いて、非リアルタイムで実行され、仮想演奏者の意図ベクトルの決定は、得られたネットワークを用いて、各拍ごとにリアルタイムで実行される。

学習アルゴリズムとして用いた、RBF(Radial Basis Function)ネットワークは、非線型関数を円形の等高線を持つ基底関数で展開する方法である。これを採用した理由は、データ間の補完能力に優れていること、ネットワークの重み算出が最小2乗法に帰着できて容易なことである。

本研究では、3人の演奏者から受け取る過去12小節の印象ベクトル(3人×12小節×3次元=108次元)を入力し、意図ベクトル(3次元)を出力するようなRBFネットワークを構成する(図6)。学習用データは、12コーラスの演奏に含まれる、過去12小節の印象ベクトルと現在の意図ベクトルとの対、N個(11コーラス×12小節×4拍=528個)である。したがって、入出力データの組は $(\mathbf{x}(k), \mathbf{y}(k))$ ($k=1, 2, \dots, N, j=1, 2, 3$)である。

ネットワークはM個($M=N$)のユニットを持ち、各ユニットの中心 \mathbf{c}_i ($i=1, 2, \dots, M$)は108次元で表される空間上の1点を表す。各ユニットは、入力ベクトル $\mathbf{x}(k)$ とユニットの中心 \mathbf{c}_i とのユークリッドノルム r を引数とするガウス関数 $\phi(r)$ を出力する(式1)。

$$\phi(r) = \exp\left(-\frac{r^2}{\sigma^2}\right) \quad 0, r > 0 \quad (1)$$

$$r = \|\mathbf{x}(k) - \mathbf{c}_i\|$$

ネットワークの出力はそれに重み $a_j(i)$ を掛け加え合わせたものである(式2)。

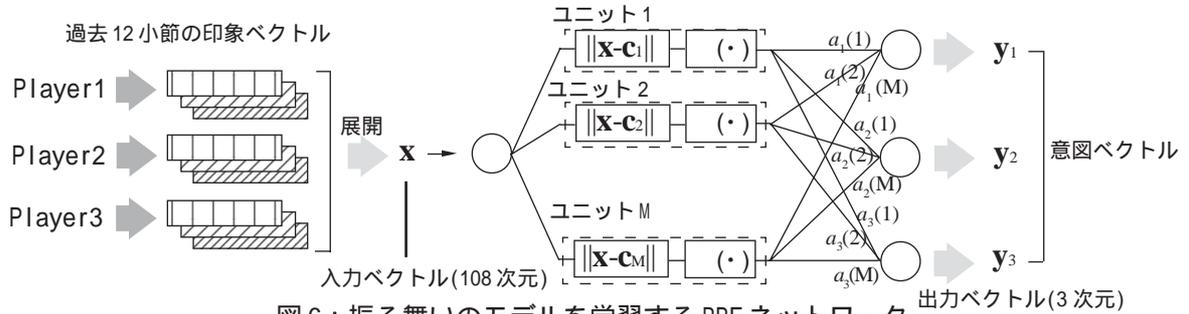


図6：振る舞いのモデルを学習するRBFネットワーク

$$f(x(k)) = \sum_{i=1}^M a_j(i) \phi(\|x(k) - c_i\|) \quad (2)$$

入力ベクトル $x(k)$ から任意の M 個をユニットの中心の候補 c_i として選択し、モデル化誤差を $e_j(k)$ とするとネットワークの入出力関係は次式で与えられる。

$$Y_j = P(x)A_j + E_j \quad (3)$$

ただし、

$$P(x) = \begin{bmatrix} p_1(1) & p_2(1) & \dots & p_M(1) \\ p_1(2) & p_2(2) & \dots & p_M(2) \\ \vdots & \vdots & \ddots & \vdots \\ p_1(N) & p_2(N) & \dots & p_M(N) \end{bmatrix}$$

$$p_i(k) = (\|x(k) - c_i\|) \quad (k=1, \dots, N)$$

$$Y_j = \begin{bmatrix} y_j(1) \\ y_j(2) \\ \vdots \\ y_j(N) \end{bmatrix}, A_j = \begin{bmatrix} a_j(1) \\ a_j(2) \\ \vdots \\ a_j(M) \end{bmatrix}, E_j = \begin{bmatrix} e_j(1) \\ e_j(2) \\ \vdots \\ e_j(N) \end{bmatrix}$$

4.1 ユニットの配置

入力ベクトル $x(k)$ から出力 $y_j(k)$ への寄与の大きいものを M 個求め、ユニットの中心 c_i として選択する。ユニットの配置は、 $y_1(k)$ 、 $y_2(k)$ 、 $y_3(k)$ それぞれについて行う。このような処理を行うと、同じ入力データが複数含まれる場合に、重みが一意に定まらないという問題を回避できるだけでなく、計算時間の短縮および汎化性の向上にもつながる。

しかし、寄与の大きいもの M 個を一度に求める方法はないため、本研究では、グラムシュミットの直交化法を用いて、寄与分が最大のものを1個ずつ選択する操作を M 回繰り返すことにより、ユニットの中心の選択を行った[8]。

グラムシュミット法を使用すると P を直交ベクトル W 、上三角行列 B に分解することができる。

$$P = WB \quad (4)$$

ここで、 $W = [w_1, w_2, \dots, w_N]$ とすると、 w_i は、 P における第 i 直交成分であり、その寄与分は、 $(w_i^T Y_j)^2 / (w_i^T w_i)$ で表される。寄与分が最大の w_i を1つ選択し、残りの直交成分について再度グラムシュミット法を繰り返すことにより、出力に対して寄与の大きいユニットが選択される。

4.2 重みの学習

誤差 E_j を最小化するような A_j を最小2乗法で推定する。式3より2乗誤差は以下で与えられる。

$$E_j^2 = Y_j - P(x)A_j^2 = (Y_j^T Y_j) - 2(Y_j^T P(x)A_j) + (P(x)A_j)^T P(x)A_j \quad (5)$$

これを A_j の各成分で偏微分すると、

$$\frac{1}{2} \frac{\partial E_j^2}{\partial A_j} = -P(x)^T Y_j + P(x)^T P(x)A_j = 0 \quad (6)$$

したがって、 A_j の推定値は式(7)で与えられる。

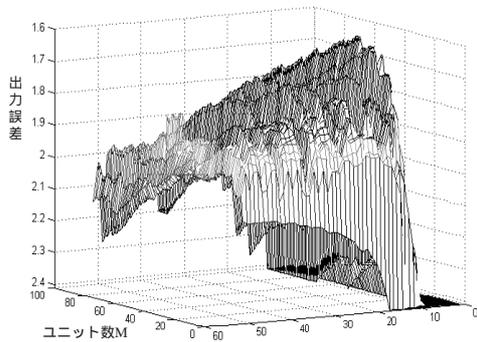
$$A_j = (P(x)^T P(x))^{-1} P(x)^T Y_j \quad (7)$$

4.3 ユニット数および σ の最適化

ユニット数および σ を最適化する。RBF ネットワークの補完能力は、ユニットの個数および σ によって大きく変化する。ここでは、この2つのパラメータを変化させ、出力誤差が最小となる値を求める。

ユニット数を1から200まで1刻み、 σ を0.5から70まで0.5刻みで、それぞれ変化させたところ、ユニット数 = 46、 $\sigma = 23.5$ が最適な値として求められた(図7)。

他のサンプルデータについても、同様の値で誤差を少なくすることができたため、演奏者のふるまいの学習を行う際には、常にこの値を用いることにした。したがって、一回の学習は、ユニットの配置および重みの学習のみで済み、所要時間は5秒以内である。



出力誤差2.0以上の部分は表示していない。
図7：出力誤差とユニット数、との関係

5 演奏生成部

演奏生成部は、あらかじめ意図空間上に配置してある演奏パターンから、仮想演奏者の意図ベクトルに合ったものを選択して出力する部分である。演奏意図空間は、演奏パターンの主観的な類似度評価に基づき構成した空間である。すなわち、似ていると感じられる演奏パターンは相互に近接し、そうでない演奏パターンは相互に離れるように配置した空間である。このような空間を用いることにより、仮想演奏者の意図ベクトルにあまり変化がない場合には、似通った演奏を出力し続けることができるし、また仮想演奏者の意図ベクトルに一致した演奏パターンがデータベース上にない場合でも、その意図ベクトルに近い演奏を出力することができる。本研究では、このような意図空間を構成するために一対比較法、および多次元尺度法(MDS)を用いている。

5.1 演奏パターンの生成

演奏パターンは、演奏者の演奏をMIDI録音したものから、8小節の長さを上限として切り出して作成した。ソロのような演奏をしている区間では、演奏的にまとまったフレーズとなるように、長さを調整して切り取った。一方、伴奏のような演奏をしている区間では、1コーラス内のあらゆる小節の頭から、区切りのよい位置までを切り取った。

5.2 一対比較実験

被験者(音楽経験者)に、77個の演奏パターンからランダムに選んだ2つの演奏パターンについて1:「似ている」、2:「少し似ている」、3:「少し似ていない」、4:「似ていない」の4段階で評定させた。その際、被験者には、あらかじめすべての演奏パターンを聴かせ、どのような演奏があるかを

把握させておいた。

5.3 多次元尺度法

Kurasukaの多次元尺度法[9]を用いて、一対比較実験により得られた類似度行列から意図空間を構成した。多次元尺度法は、2つの演奏パターン j と演奏パターン k の類似度を d_{jk} 、多次元空間での距離を d_{jk} とすると、類似度の高い演奏フレーズほど距離が近くなるように多次元空間内の点の位置を決定するものである(式5)

$$d_{jk} > d_{lm} \text{ ならば } d_{jk} < d_{lm} \quad (5)$$

このとき、式5が成立する度合いは、ストレス値 S で表される。意図空間は次元数とストレス値 S の両方が小さくなることが望ましい。そこで、1次元から5次元の各次元数におけるストレス値を計算した。その結果、1、2次元ではストレス値が高かったが、3次元以降では、ストレス値が十分小さくなったため、意図空間の次元数は3次元とした。

5.4 正準相関分析

正準相関分析を用いて、意図ベクトルと出力演奏の物理特徴ベクトルとの相関を求めた。用いた物理量は3.2(1)のものと同様である。分析の結果、正準相関の値が0.88となった。これにより、意図ベクトルから出力演奏の物理特徴ベクトルへの写像およびその逆写像が可能となった。逆写像は、演奏者の振る舞いのモデルを学習するとき、新たな演奏パターンをデータベース上加えるときに用いる。

5.5 出力演奏の生成

出力演奏は、データベースにある100個の演奏パターンから選択されたものを切れ目なくつなげ合わせることで生成される。仮想演奏者は、前の演奏パターンの終わりになると、コード進行の制約を満たした演奏パターンの中から意図ベクトルに最も近いものを次の演奏パターンとして選択する。

6 実験結果

システムをパソコン(Pentium 650MHz, Windows98)上に実装した。3人の人間が、クリック音にあわせて行ったセッションのMIDI演奏記録を用いて学習を行ったところ、それぞれの演奏者の振

る舞いのモデルが獲得できた。得られた振る舞いのモデルを用いて仮想演奏者を動作させたところ、人間の演奏者1人と仮想演奏者2人、人間の演奏者2人と仮想演奏者1人、仮想演奏者3人など、様々な構成でセッションを行うことが可能であった。演奏はシステムが出すカウント音によって開始する。システムを用いて人間の演奏者がセッションを行うごとに、新たな演奏者の振る舞いのモデルが獲得できた。

仮想演奏者が振る舞いのモデルに基づき動作していることを確認するため、人間Cと仮想演奏者A,B(仮想演奏者Bは、人間Cの振る舞いのモデルをもつ)でセッションを行った。

その結果、人間の演奏者Cと仮想演奏者Bとは、ソロに入るタイミングが一致したり、類似したフレーズを演奏する傾向があった。図8は、特に類似していると感じられた部分4コーラスでの3者の演奏意図の変化を表している。12コーラス分の演奏の各時刻における人間の演奏者の意図ベクトルと仮想演奏者の意図ベクトルとの距離を合計し、その平均値を比較したところ、人間の演奏者Cと仮想演奏者Aでは、平均9.33と離れていたが、演奏者Cと仮想演奏者Bでは、平均3.56と近かった。

以上の実験により、仮想演奏者が演奏者の振る舞いのモデルに基づいた振る舞いをしていることが確認された。

7 まとめと今後の課題

本稿では、人間の演奏者の振る舞いを模倣した仮想演奏者と人間の演奏者がインタラクションできるようなジャムセッションシステムについて述べた。本システムでは、模倣しようとしている演奏者の入出力関係を、統計的に学習することにより、演奏者の振る舞いのモデルを獲得することを可能にした。

本システムでは、仮想演奏者の振る舞いのモデルを過去12小節の印象ベクトルの履歴によって学習したが、今後、より大局的に見た時系列の変化も考慮にいった学習アルゴリズムについて検討していく予定である。

謝辞

本研究を開始する機会を与えて頂いた中央大学理工学部教授(電子技術総合研究所 ヒューマンメディアグループ ラボリーダー)加藤俊一氏に深謝いたします。実験に協力して頂いた、斉田康広氏、千田真一氏、橋本大輔氏に感謝いたします。

参考文献

- [1] M. Nishijima and K. Watanabe: Interactive music composer based on neural networks, *Proc. of ICMC*, pp.53-56, 1992.
- [2] 青野裕司, 片寄晴弘, 井口征士: バンドライクな音楽アシスタントシステムについて, *情報処理学会研究報告*, 94-MUS-8, Vol.94, No.103, pp.45-50, 1994.
- [3] 青野裕司, 片寄晴弘, 井口征士: アコースティック楽器を用いたセッションシステムの開発, *電子情報通信学会論文誌*, D- , Vol. J82-D- , No.11, pp.1847-1856, 1999.
- [4] 和気早苗, 加藤博一, 才脇直樹, 井口征士: テンションパラメータを用いた協調型演奏システム - JASPER -, *情報処理学会論文誌*, Vol.35, No.7, pp.1469-1481, 1994.
- [5] 近藤欣也, 片寄晴弘, 井口征士: 音楽情報から奏者の意図を理解する伴奏システム JASPER++, *情報処理学会第46回全国大会*, 1-373, 7Q-8, 1993.
- [6] 後藤真孝, 日高伊佐夫, 松本英明, 黒田洋介, 村岡洋一: 仮想ジャズセッションシステム: VirJa Session, *情報処理学会論文誌*, Vol.40, No4, pp.1910-1921, 1999.
- [7] 日高伊佐夫, 後藤真孝, 村岡洋一: すべてのプレーヤーが対等なジャズセッションシステム, ベーシストとドラマーの実現, *情報処理学会研究報告*, 96-MUS-14, Vol.96, No.19, pp.29-36, 1996.
- [8] S.Chen, C.F.N. Cowan and P.M. Great: Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks, *IEEE Transactions on Neural Networks*, Vol.4, pp.246-257, 1991.
- [9] J.B. Kruskal and M. Wish, *Multidimensional Scaling*, Sage Publications, 1978.

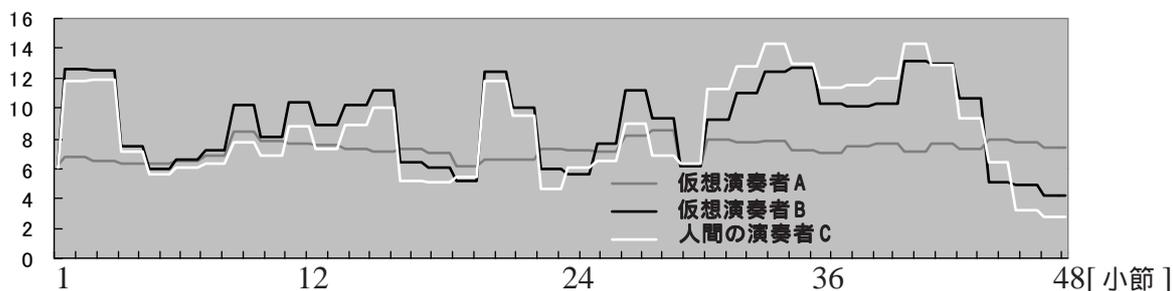


図8: 3人の演奏者の演奏意図の比較(3軸の意図ベクトルのうち1成分のみを表示)