

演奏者の個性を模倣する
セッションシステムに関する研究

工学研究科

筑波大学

2003 年 3 月

浜中 雅俊

STUDY ON A SESSION SYSTEM
IMITATING MUSICIAN'S
PERSONALITY

Doctral Program in Engineering

University of Tsukuba

Date(2003, March)

Masatoshi Hamanaka

目次

第1章 序論	11
1.1 研究の背景	11
1.2 関連研究	13
1.2.1 演奏レンダリングシステム	13
1.2.2 ジャムセッションシステム	15
1.2.3 関連研究に対する議論のまとめ	18
1.3 研究の位置づけ	18
1.4 本論文の構成	20
第2章 演奏者の個性の模倣	25
2.1 はじめに	25
2.2 システムの全体像	26
2.3 演奏者の振る舞いの模倣	27
2.3.1 印象空間の導入	29
2.3.2 意図空間の導入	29
2.3.3 振る舞いのモデルの獲得	30
2.4 演奏者固有のフレーズの模倣	30
2.4.1 発音時刻のクオンタイズ	31
2.4.2 フレーズの分割	32
2.4.3 フレーズの再利用	32
2.5 発音時刻のゆらぎの模倣	32
2.5.1 ゆらぎの個性の獲得	33
2.5.2 逆関数法によるゆらぎの付加	33
2.6 まとめ	34

第3章	演奏者の振る舞いの模倣	35
3.1	はじめに	35
3.2	振る舞いのモデルの学習	35
3.2.1	印象ベクトルの算出	36
3.2.2	意図ベクトルの算出	36
3.2.3	振る舞いのモデルの獲得	37
3.3	仮想演奏者の構成	37
3.3.1	印象算出部	37
3.3.2	意図決定部	42
3.3.3	演奏生成部	46
3.4	実験結果	49
3.5	まとめ	51
第4章	演奏者固有のフレーズの模倣	53
4.1	はじめに	53
4.2	発音時刻のクォンタイズ	54
4.2.1	学習に基づくクォンタイズ	55
4.2.2	最適発音時刻系列の推定	61
4.2.3	モデルパラメータの学習	64
4.2.4	クォンタイズ結果	71
4.3	フレーズ分割	77
4.3.1	Generative Theory of Tonal Music	78
4.3.2	GTTM 以外のグルーピング手法	82
4.3.3	ボロノイ線図に基づくグルーピング	83
4.3.4	ボロノイ線図に基づくフレーズ分割	89
4.4	フレーズの再利用	90
4.4.1	意図空間の再構成	90
4.4.2	局所自己相関関数を用いた特徴の抽出	93
4.4.3	正準相関分析	95
4.4.4	出力演奏の生成	95
4.5	まとめ	96

第5章	発音時刻ゆらぎの模倣	99
5.1	はじめに	99
5.2	発音時刻ゆらぎの順モデル	101
5.3	逆関数法による確率分布乱数	102
5.4	発音時刻ゆらぎの付加	103
5.5	ゆらぎの付加に関する考察	105
5.6	まとめ	106
第6章	評価と考察	107
6.1	はじめに	107
6.2	定量的な評価	107
6.3	心理実験による評価	109
6.4	実験結果・考察	110
6.5	まとめ	111
第7章	結論	113
	謝辞	117
	参考文献	127
付録A	主成分分析による印象語選択	133
付録B	特徴量抽出のための予備実験	137
付録C	正準相関分析	141
付録D	グラムシュミットの直交化法	145
付録E	スタインのパラドックス (経験的ベイズ法)	147
付録F	MDA-OR(最小次元解析)	149

目 次

1.1	演奏レンドリングシステムの構成 (文献 [平賀 02] より引用)	14
1.2	論文の構成	23
2.1	演奏者の3種類の個性	26
2.2	システムの全体像	27
2.3	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	28
2.4	演奏者の振る舞い	29
2.5	演奏者の振る舞いのモデル	30
3.1	振る舞いのモデルの学習の手順	36
3.2	仮想演奏者の構成	38
3.3	印象語の評定に用いたアプリケーション	40
3.4	様々な物理特徴量の抽出	41
3.5	被験者とシステムが示した存在感の値の比較	42
3.6	振る舞いのモデルを学習する RBF ネットワーク	43
3.7	ユニット数 M と σ を変化させたときのモデル化誤差の値	46
3.8	演奏パターンの生成	47
3.9	システムの出力画面	49
3.10	3人の演奏者の意図ベクトルの変化 (3軸の意図ベクトルのうち1成分のみを表示)	50
4.1	クオンタイズにおける順モデルと逆モデル	56
4.2	1拍の隠れマルコフモデルの概略図	59
4.3	1つの HMM によるモデルと4種類の HMM に分岐したモデル	60
4.4	最尤な状態遷移を求めるトレリス	62
4.5	付点2分音符の状態遷移の例	63
4.6	4つの HMM による探索木の例	64

4.7	人工, 実演奏データから求めた $b_5(k)$	66
4.8	演奏 C2 から求めた a_{ij}	67
4.9	教師つき/なし学習で求めた $b_4(k)$	71
4.10	クオンタイズ結果の例	74
4.11	グルーピングの例	77
4.12	正しい/誤ったグルーピングの例	79
4.13	GTTM とベン図によるグルーピングの表現	83
4.14	n 個の点によるボロノイ線図	84
4.15	水平な 2 直線によるボロノイ線図	85
4.16	2 本以上の直線によるボロノイ線図	86
4.17	ボロノイ線図に基づくグルーピング結果	87
4.18	GTTM に基づき手作業によるグルーピング結果	87
4.19	意図空間と出力演奏の物理的な特徴との相関	91
4.20	Kruskal と MDA-OR による空間配置	92
4.21	時間軸を伸縮させたピアノロール	93
4.22	局所領域の変位パターン	94
4.23	フレーズの接続	96
5.1	発音時刻ゆらぎの順モデル	101
5.2	実演奏データから求めた $b_1(k); b_5(k); b_9(k)$	102
5.3	標準正規分布とその累積分布関数	103
5.4	和音へ同じ確率分布でゆらぎを付加した例	104
5.5	3 和音の場合における分布の分割	104
5.6	和音の 1 音目と 2 音目以降の $b_9(k)$	105
5.7	和音の発音時刻間隔 IOI の分布	106
6.1	評定結果と模倣した演奏者が一致した割合	110
B.1	ソロと伴奏でのベロシティーの比較	137
B.2	ソロと伴奏での音高の比較	138
B.3	ソロと伴奏でのコードの主音からの度数の比較	138
B.4	ソロと伴奏でのピッチベンドの比較	139

E.1 10人の野球選手に対するジェームズ・スタイン推定量	148
-------------------------------	-----

表 目 次

1.1	人工知能と人工感性，人工個性	12
4.1	正解データに含まれる4種類の拍の割合	75
4.2	機械クオンタイズと本手法の比較（音単位の一一致率）	75
4.3	1つのHMMによるモデルと4つのHMMによるモデルの比較（拍単 位の一一致率）	75
4.4	他の演奏のモデルパラメータでのクオンタイズ結果	75
4.5	3つのスケーリング比におけるグルーピング結果	88
4.6	本手法と小節線での分割結果の比較	90
6.1	心理実験の結果	112
A.1	心理実験から得られたデータ	133
A.2	固有値と累積寄与率	135
A.3	存在感，躍動感，重厚感の因子負荷量	135
C.1	印象ベクトルと物理特徴ベクトルのデータ	141

第1章 序論

科学技術の発達に伴い，工学的手法を用いて人間の様々な能力を計算機上で実現しようとする試みが進められている．このような試みは，その成果の実応用が期待されるばかりでなく，人間の仕組みを解明する上でも重要である．本論文では，計算機上の仮想演奏者と人間の演奏者とが即興演奏により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論じる．音楽を通した人間と人間とのコミュニケーションの中でも，特にジャムセッションは，コード進行など最小限の約束ごとのみを決め，順次ソロパートを交代しながら演奏していくため，ある演奏者の演奏がそこで演奏している演奏者全員に影響を与え，他の演奏者の新たな演奏となって返ってくるというインタラクションが頻繁に起こる．また，ジャムセッションでは演奏者に固定した主従関係が存在せず，すべての演奏者が対等な立場でインタラクションできる．したがって，そこでのインタラクションの仕方は，演奏者の個性によって大きく異なっている．このようなジャムセッションにおいて，実在する人間と同じようにインタラクションが可能な仮想演奏者を実現するためには，これまで取り扱いが困難とされてきた人間の個性や感性を工学的にモデル化する必要がある．

1.1 研究の背景

人間の能力の計算機上での実現は，まず，人工知能 [RN95] の研究分野で試みられた．そこでの目的は，あたかも人間のように認識や判断などをする計算機システムの実現である．実際に実用化されたものには，パターン認識 [大津 81] による人間の知覚の代行や，エキスパートシステムによる人間の判断の代行等があるが，そのようなシステムでは客観性が求められるため，感性や個性など主観的な要因は排除されてきた．

一方，感性工学の分野では，これまで，感覚器でのセンシングから心理的認知的現象としてのイメージまでの過程の工学的モデル化が進められており，その過程にお

表 1.1: 人工知能と人工感性，人工個性

領域	対象	目的	実用化用途
人工知能 (artificial intelligence)	知能 (知識)	知覚，判断，思考，学習，記憶 などの過程の計算機上での実現	パターン認識 エキスパートシステム
人工感性 (artificial Kansei)	感性	人間と計算機とのよりよいイン タフェイスの実現	印象語による商品検索 デザイン支援システム
人工個性 (artificial personality)	個性	各人の固有の癖の獲得とその計 算機上での再現	仮想音楽家 仮想デザイナー

ける個別性について検討されてきた．そこでの主な目的は，人間と計算機とのよりよいインタフェイスを実現するために，計算機上に人工感性を構築することである．実際に実用化されたものとしては，類似検索 [加藤 89, 鈴木 99] や形容詞（印象語）検索 [栗田 92, 加藤 92] による商品検索システム等がある．しかし，そこで扱われた個別性とは，商品の物理的特徴と検索空間とのマッピングの個人差のみであった．

あたかも実在する人間のような個性を持った計算機システムを構築するためには，人工知能や人工感性を実現するだけでなく，各人の固有の癖をなんらかの形で獲得し，それを用いて人工的に個性を再現する必要がある．ここではそれを，人工個性 (artificial personality) と呼ぶことにする．表 1.1 に人工知能と人工感性，人工個性のそれぞれの対象，目的，用途についてまとめる．

ここで注意すべきは，感性や個性という用語には様々な意味があり，しばしば混同されがちなことである．本論文では両者を以下のような意味で用いることにする．まず，感性とは，環境や他人など外界からの情報入力によって触発される心の動きに関わるもので，外部からの情報を自分自身のフィルタを通じで認知する機能である．一方，個性とは，外界からの入力に対して出力を決める過程における考え方や行動の癖の個体差のことである．したがって，人工個性を実現するためには，外界との情報の入出力をより豊かにするという人工感性の考え方と，入力と出力を定式化し計算機モデルを構築するという人工知能の考え方の両方が必要である．

人工個性の技術の最終的な目標は，実在する演奏者の個性を獲得し，それを模倣した仮想演奏者を計算機上で実現したり，工業デザイナーの個性を模倣した仮想デザイナーを実現することである．もしそれが実現すれば，獲得した個性をネットワーク上で配信し，一般のユーザがそれを利用するということが可能となる．また，専

門家が作品や製品を「創造」し、ユーザがそれを「鑑賞」したり「使用」するという従来の枠組みに対して、ユーザが制作に「参加」という新たな枠組みを提供することが可能となる。例えば、音楽の分野では、著名な作曲家の個性をもった仮想作曲家を生成し、自分好みの曲を作曲させることや、ある演奏家の個性をもった仮想演奏者を生成し、それと共演すること等も可能となる。このような研究は、音楽初心者にも音楽を楽しませるアミューズメント性をもつだけでなく、プロの演奏家にとっても、音楽制作の生産性をあげる技術の一つとして期待がもたれる。具体的には、ギターしか弾けない演奏者は、ギター以外のパートを、様々な個性をもった仮想演奏者に演奏させることにより、バラエティに富んだ曲を短時間で制作できるようになる。

本研究では、上記のような様々な効果が期待されている音楽分野における人工個性に着目する。そして、音楽分野における人工個性実現の第一歩として、ジャムセッションにおける演奏者の個性を扱う。本論文では、人間の演奏者の個性を模倣した仮想演奏者と人間の演奏者がジャムセッションできるようなシステムを実現するための手法について述べる。

1.2 関連研究

演奏者の個性に関する研究事例として、演奏レンダリングシステムの研究とジャムセッションシステムの研究があげられる。ここでは、それぞれのシステムの従来の研究、開発の流れについて整理する。

1.2.1 演奏レンダリングシステム

演奏レンダリングシステムは、演奏者が楽譜を使って演奏する過程を計算機上で実現しようとするシステムである。その主な対象は、有名なクラシック曲など楽譜とそれに対応した演奏記録が存在する曲である。具体的には、楽譜情報（MIDI¹ファ

¹Musical Instrument Digital Interface の略。電子楽器の演奏情報を送受信するための共通の規格。MIDI 規格では、楽器の様々な演奏動作に関して各 MIDI メッセージが定義されている。主な MIDI メッセージには、ノートオン（ペロシティー、音高を指定して、音を発音させる）、ノートオフ（発音した音を止める）、ピッチベンドチェンジ（発音中の音のピッチを相対的に上下にずらす）などがある。

イル等)を入力として、人工知能や統計処理の様々な手法を用いて、人間のように情緒や抑揚など表情のついた演奏を出力するものである。一般に、演奏レンダリングシステムは、図1.1(文献[平賀02]より引用)のような構成をしている。ステージ1では、楽譜情報とそれに対応する実演奏記録を与えて、分析や学習を行い、レンダリングに必要な情報を事例やルールとして取り出す。ステージ2では、レンダリングすべき課題曲の楽譜情報を入力し、演奏データを出力する。マニュアルと書いてあるのは、人手によって処理や修正や補足を行うことを表す。特に、ステージ1や2の処理を行わず、ステージ3のマニュアルだけで出力を生成する場合を打ち込みと呼ぶ。レンダリングシステムの評価法としては、コンテスト形式が試みられている[平賀02, HHH+02, 片寄02a, 片寄02b, 橋田02]。現在のところ、芸術的によりよい演奏を作るという観点では、レンダリングエンジンを使った演奏より、打ち込みによる演奏のほうが優れているという状況にあり、音楽的観点とシステムの観点を独立に評価するという方法がとられている。

演奏レンダリングシステムは、楽譜が存在する曲を主な対象としており、ジャムセッションの即興演奏はこれまで対象外とされてきた。しかし、ジャムセッションで同じフレーズを複数の演奏者が弾いた場合でも、演奏者により発音時刻のゆらぎなどが生じる。演奏者の発音時刻ゆらぎの個性を模倣する手法については5章で取り上げる。

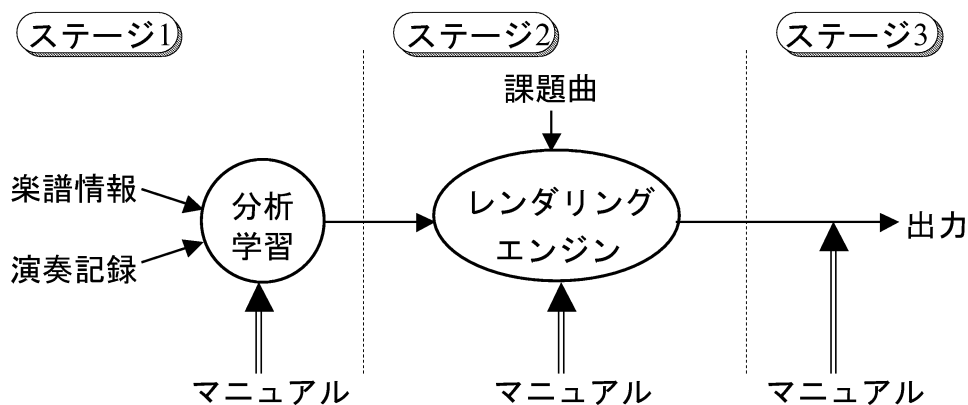


図 1.1: 演奏レンダリングシステムの構成 (文献 [平賀 02] より引用)

1.2.2 ジャムセッションシステム

ジャムセッションシステムは、仮想演奏者と人間の演奏者がお互いに影響を及ぼしあいながら、即興演奏していくシステムである。ジャムセッションシステムの研究には、様々なアプローチがあるが、仮想演奏者と人間の演奏者のインタラクションをより豊かなものにすることが共通の課題である。以下、これまで報告されたジャムセッションシステムの特徴についてまとめる。

2 個性を考慮していないジャムセッションシステム

従来のジャムセッションシステム [AKI95, 青野 99, PSR93, SP93] 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。

青野らが開発した Band Master [AKI95], AdliBand [青野 99] は、あらかじめ楽譜を必要としないジャムセッションを可能としていた点が特徴である。システムは、セッション中で入力された演奏から、コード進行やリズムパターン、ベースラインなどを抽出し、抽出したパターンを再合成して楽曲を構成する。さらに AdliBand では、多くのジャムセッションシステムが、システムの入出力に MIDI を用いているのに対し、入力に音楽音響信号を用い、リアルタイムでの和音認識を行った。

Pennycook らが開発した即興演奏のための実時間解析システム [PSR93, SP93] は、入力演奏を 2 小節から 10 小節の長さのセグメントに分割し、そのセグメントから特徴を抽出するとともにデータベース化し、セグメントを演奏の生成に再利用しようとしていた点が特徴である。しかし、実際に実装されたのは、聴取部分のみで、具体的な再利用の手法については提案されていなかった。セグメントを再利用するためにデータベース化するという考え方自体は、4 章における、フレーズ分割・再利用の考え方と一致する。Pennycook らのセグメント分割手法については、4 章で再度取り上げる。

2 個性パラメータをもつジャムセッションシステム

文献 [Row92, 和気 94, 近藤 93, 金森 95, 後藤 99, 日高 96, GHM⁺96] のシステムでは、パラメータやルールの調整により、異なる個性の仮想演奏者のモデルを設定することはできても、実在する人間の演奏者の個性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。

文献 [Row92] で発表された Cypher は、聴取部と演奏部および、聴取部と演奏部の関係をユーザが設定するための GUI からなるシステムであった。聴取部では、音の密度、音域、強弱や、コード、フレーズなどを認識する。また、演奏部では、あらかじめ用意した演奏やアルゴリズム作曲した演奏を出力したり、入力された演奏に様々な変形処理を行い出力することができる。そして、あらかじめユーザが設定したような演奏が聴取部に入力されると、演奏部ではそれに対応した処理を選択し演奏を出力する。Cypher は、ユーザの設定により仮想演奏者に個性を与えるシステムの一例であるが、あらかじめ仮想演奏者の動作を細かく設定しておく必要があることから、ジャムセッションシステムというよりもインタラクティブ作曲のための支援システムとしての意味合いが強かった。

文献 [和気 94, 近藤 93] で発表された Jasper は、12 小節 1 コーラスの典型的なブルース進行の曲を対象とし、ピアノ演奏者のソロに対して、計算機がドラム、ベースの伴奏を行うシステムであった。具体的には、まず、ピアノ演奏の音域、強弱、音数から、演奏者の盛り上がりに対応するテンションパラメータを計算し、次小節でのテンションパラメータを決定する。そして、そのテンションパラメータに対応した長さ 1 小節の演奏パターンを出力する。前の小節のテンションパラメータから、次の小節のテンションパラメータを決定するためのルールを、個性データベースに記述することにより仮想演奏者の個性を設定することを可能としていた。

文献 [金森 95] では、他の演奏者の演奏を認識する聴取部と、自らのパートを演奏する生成部とで仮想演奏者を構成し、聴取部と生成部それぞれに個性プロフィールを与えることを提案していたが、実際に実装されていたのは聴取部のみで、演奏を生成するしくみについては、提案されていなかった。聴取部では、入力演奏に対応した印象語を音楽家の経験則に基づき出力する。これは、計算機上で音楽家の人工感性を実現するための試みと考えることができる。個性プロフィールは、その過程における入力演奏と印象語との対応関係を定義したものであり、感性を設定するためのパラメータであると考えられる。

文献 [後藤 99, 日高 96, GHM⁺96] で発表された VirJa Session は、システム外

部から調節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。VirJa Session の特徴は、すべての演奏者が固定した主従関係をもたず、対等な立場でのジャムセッションを実現した点である。各演奏者は LAN を介して分散環境で実装されており、演奏者は相手が、人間であるか仮想演奏者であるかは問わない。また、仮想演奏者にカメラによる視覚を与えるとともに、CG により仮想演奏者を可視化し、ジェスチャーによるインタラクションも可能とした。

2 学習機能をもつジャムセッションシステム

演奏者の振る舞いを模倣するためには、その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、すなわち、演奏者の入出力関係を学習により獲得する必要がある。文献 [NK89, NW92] で発表された、Neuro-Musician では、あらかじめ用意した長さ 8 小節の入出力演奏パターン 30 対をニューラルネットで学習することにより、人間の 8 小節の演奏に対して、8 小節の演奏パターンを出力するシステムを実現した。この研究は、演奏者のモデルを学習により獲得しようとした点が優れており、複数の演奏パターンの補間により新たな演奏パターンが生成できることを示していた。しかし、そのシステムは、人間の演奏者と仮想演奏者が交互にソロを弾くという制限された条件下でのみ動作するもので、人間と仮想演奏者が同時にソロや伴奏を弾くことはできなかった。また、モデルを学習する際にも、8 小節の入出力データをあらかじめ用意しなければならないという問題があった。実際のジャムセッションの演奏では、相手の 8 小節のフレーズに対して、必ずしも 8 小節のフレーズで反応しているとは限らず、相手のどのフレーズに対して、どのフレーズで反応しているかは明らかではない。したがって、文献 [NK89, NW92] の手法を用いて、ジャムセッションの演奏記録から演奏者モデルを獲得することは困難であった。

1.2.3 関連研究に対する議論のまとめ

従来の演奏レンドリングシステムとジャムセッションシステムに対する本節の議論を以下にまとめる。

2 演奏レンドリングシステム

有名なクラシックの曲が対象で、これまでジャムセッションなどの即興演奏は対象外であった。

2 個性を考慮していないジャムセッションシステム

人間の演奏者に追従した演奏を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。

2 個性パラメータをもつジャムセッションシステム

パラメータやルールの調節により、異なる個性の仮想演奏者のモデルを設定することはできても、実在する人間の演奏者の個性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。

2 学習機能をもつジャムセッションシステム

従来の学習機能をもつジャムセッションシステムは、限定された条件下のみで動作するもので、ジャムセッションの演奏記録から演奏者モデルを獲得することは困難であった。

1.3 研究の位置づけ

本研究では、以上の背景のもと、実在する人間の演奏者の個性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を実現する。

本研究の主な意義は以下の3点である。

1. ジャムセッションシステムの研究を通して、人工個性の実現を試みたこと。
2. ジャムセッションシステムの演奏記録から演奏者モデルの獲得を実現したこと。
3. ジャムセッションシステムの各要素技術が高い性能を持ち、他の音楽システムでも利用可能なこと。

第一の意義は、計算機上に人間の個性を模倣した仮想演奏者を実現した点である。これは、各人の癖を獲得しそれを人工的に再現する人工個性実現の試みの一つである。本研究では人工個性の対象として、ジャムセッションの即興演奏における演奏者の個性を扱った。その理由は、ジャムセッションでは演奏者の個性の違いが演奏に大きな影響を及ぼすからである。ジャムセッションシステムという題材を通じて、人間の個性の工学的な扱いを確立することが、本研究の最大の目的である。

第二の意義は、演奏記録の統計的学習により演奏者モデルを獲得した点である。ジャムセッションシステムの研究の一つの目標は、いかに人間らしい個性を持った仮想演奏者を生成するかである。本研究では、演奏記録に演奏者の個性が反映されていると考え、統計的な学習によりモデル化を行った。このような演奏記録からの学習という考え方は、今後のジャムセッションの研究に一つの方向を与えることが期待される。

演奏者モデルの獲得を実現するためには、統計的手法が重要な役割を担っている。たとえば、ジャムセッションの演奏から重要な特徴量を抽出する際には、パターン認識および多変量解析が、基本的な手法として重要である。また、演奏者の入出力関係など非線形な対象をモデル化するためには、RBF ネットワークのような手法が有効である。さらに、演奏の発音時刻や、そのゆらぎなどを扱うためには、隠れマルコフモデルのような確率モデルが有効である。上記のような統計的手法を適切に応用していくことによって、演奏者の様々な個性をモデル化していくことが可能となる。

第三の意義は、発音時刻のクオンタイズや、フレーズ分割など、本ジャムセッションシステムを構成する各技術が、音楽情報科学の研究分野において重要な技術であり、今後、他の音楽システムで利用していくことが期待される点である。本研究で実装したすべての要素は、人間の手作業によるパラメータの設定などが不要で自動で動作する。これは、他のシステムなどで利用する際には、大きな利点となる。

1.4 本論文の構成

本論文は本章も含めて7つの章からなる。

第2章では、まず、ジャムセッションにおける個性として、演奏者の振る舞いの個性、演奏者固有のフレーズの個性、演奏の発音時刻ゆらぎの個性の3種類について述べ、システムの全体像について説明する。そして、演奏者の個性を獲得してシステムを実現する上での問題点についてまとめる。

システムは、演奏者の個性を収めた3種類の個性データベースと、その個性を用いて動作する仮想演奏者からなる。演奏者の振る舞いの個性は、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獲得することができる。このような学習は、入力演奏からその演奏者の出力演奏を求める関数を獲得す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り、限られた長さの演奏記録からモデルを獲得することが困難である。そこで、心理実験の結果を、主成分分析、多次元尺度法、正準相関分析を用いて分析し、入力を抽象化した印象空間と、出力を抽象化した意図空間を導入することにより、入出力関係の学習を可能とする。演奏者固有のフレーズの個性は、模倣しようとしている演奏者の演奏記録から、フレーズを切り出して再利用することにより模倣することができる。しかし、フレーズ切り出しの際ゆらぎを除去せずに行うと、フレーズの接続時につなぎ目が不自然となる。そこで、確率モデルを用いて発音時刻のゆらぎを除去した後にフレーズ切り出しを行い、演奏の再利用を可能とする。発音時刻ゆらぎの個性は、発音時刻の特徴・癖を演奏記録から確率モデルとして学習することにより獲得できる。そして、獲得した確率モデルに基づくゆらぎを付加することにより、発音時刻ゆらぎの模倣を可能とする。

第3章では、演奏者の振る舞いを学習する手法について述べ、実際に演奏記録から振る舞いのモデルを学習する。そして、獲得した振る舞いのモデルに基づき仮想演奏者が動作することを確認する。

振る舞いのモデルを学習するためには、まず、演奏者が聴取した演奏の印象空間上での位置(印象ベクトル)を算出する。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を座標軸に持つ空間である。次に、演奏者が出力し

た演奏から，その演奏を生成するような意図空間上での位置（意図ベクトル）を算出する．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，被験者実験の結果求めた主観的類似度（被験者の主観に基づく類似度）に基づき配置した空間で，どのような演奏を生成したいかを表す．最後に，RBF ネットワークを用いて，印象ベクトルから意図ベクトルへの非線形関数を学習することにより振る舞いのモデルが獲得できる．そして，演奏者とその演奏者から獲得した振る舞いのモデルを用いて動作する仮想演奏者でジャムセッションし両者を比較することにより，演奏者の振る舞いが模倣できているか確認する．

第4章では，演奏者固有のフレーズを模倣するための手法として，発音時刻のクォンタイズ，演奏のフレーズ分割，演奏の再利用法について説明する．

発音時刻のクォンタイズは，ゆらぎを含むジャムセッションの演奏の発音時刻から，演奏者が元々弾こうとした正規化された発音時刻を求める手法である．ここでは，実際の演奏から，正規の発音時刻を推定する問題を，発音時刻ゆらぎが生じる問題の逆問題ととらえ，確率モデル（隠れマルコフモデル）を用いて定式化する．そして，確率モデルのモデルパラメータを演奏記録から統計的に学習することによって，各演奏に適した確率モデルを使ったクォンタイズを可能とする．モデルパラメータの教師なし学習には，Baum-Welch アルゴリズムとヘルドアウト補間法を用いる．

演奏のフレーズ分割は，音楽の階層的な構造を考慮して，その切れ目となる部分で自動で分割する手法である．楽曲のグルーピングに関する理論としては，Generative Theory of Tonal Music (GTTM) があるが，GTTMのグルーピング構造分析は，単旋律の楽曲が対象で，和声に関する分析が欠如しているため，我々が目標とするジャムセッションの演奏のグルーピングには対応していない．また，ルールが定量的に定義されておらず，複数のルール間の順位が不明確で競合が起きるなど，理論を計算機上へ実装することは困難だと考えられている．そこで，GTTMを直接計算機上へ実装するのではなく，GTTMによるグルーピング結果と近い結果を出すボロノイ線図を用いたグルーピング手法を提案し，フレーズ分割を行う．

演奏の再利用法は，切り出したフレーズを再利用して，仮想演奏者が演奏を生成するための手法である．そのためには，仮想演奏者がどのような意図のときにそのフレーズを使用すれば良いか分からなければならない．そこで，各フレーズに対応する意図ベクトルを求め，意図空間上に配置する．具体的にはまず，多変量解析手

法の MDA-OR を用いてフレーズを空間配置し，局所自己相関関数によりフレーズの特徴抽出を行う．そして正準相関分析により，意図ベクトルとフレーズ特徴との関連づけを行い，仮想演奏者のそのときの意図ベクトルに対応したフレーズを次々と接続していくことで演奏を生成する．

第5章では，発音時刻ゆらぎを模倣する手法について述べ，ゆらぎのない演奏にゆらぎの付加を行う．具体的には，クオンタイズのために，演奏記録から学習した確率モデルを用いて，演奏にゆらぎを付加する．その際，和音の構成音のそれぞれに，同じ確率分布を用いてゆらぎを付加してしまうと，ゆらぎ付加前の和音の構成音の発音順序と，付加後の発音順序が異なってしまうという問題が起きる．そこで，和音の構成音の数で，イベント時刻方向に等分割した分布をつくり，その分布に基づくゆらぎを付加することで，和音の構成音の順序が変化しないようにする．

第6章では，構築したシステムの評価を行う．ジャムセッションシステムには，まだ絶対的な評価の枠組みが確立されていない．そこで，定量的な評価と心理実験による評価の両方でシステムを評価する．そして，最終的にはチューリングテストからヒントを得た心理実験を行い，仮想演奏者が演奏者の個性を模倣できているか確認する．

第7章では，結論およびジャムセッションシステムの今後の行方について考察する．

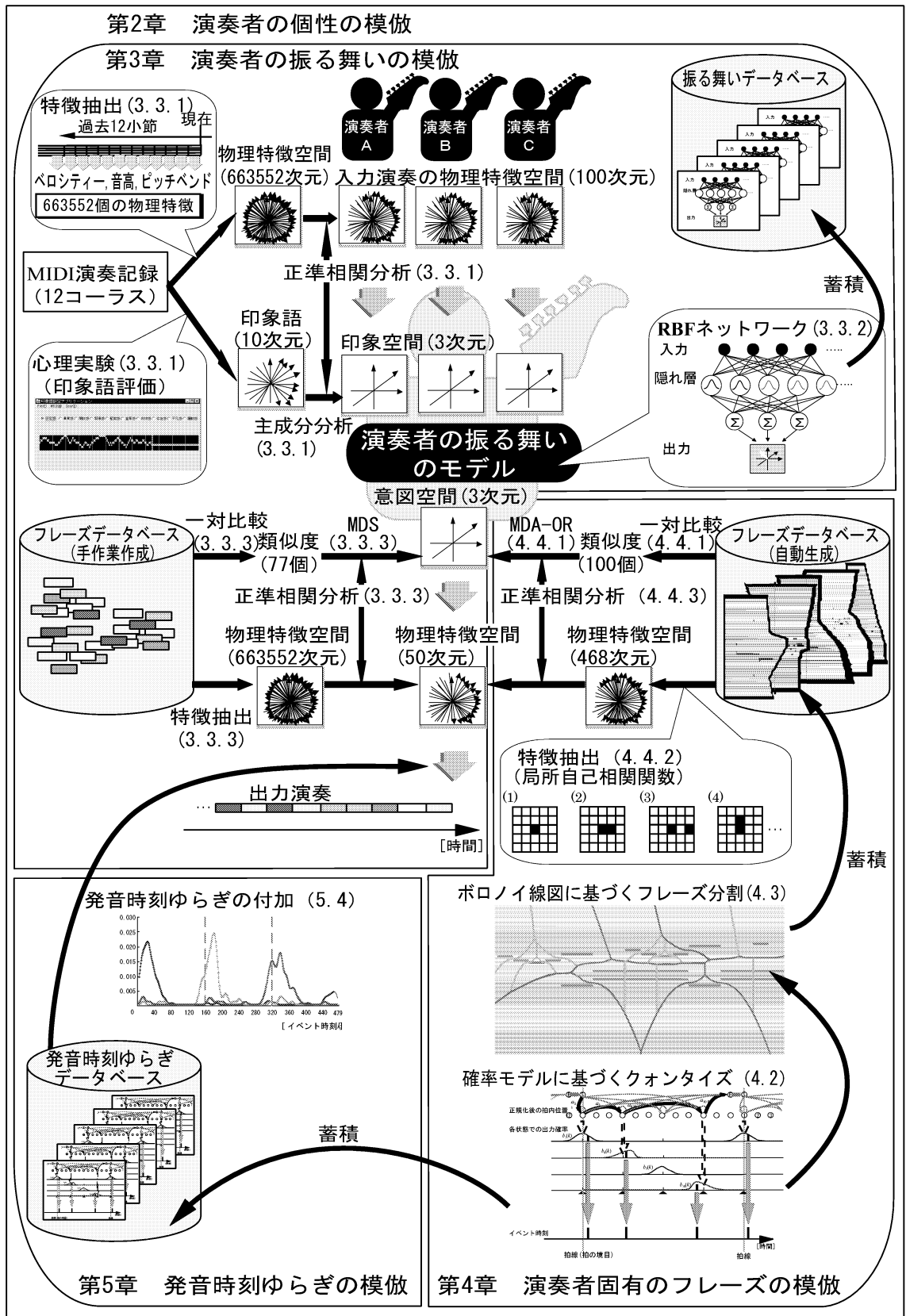


図 1.2: 論文の構成

第2章 演奏者の個性の模倣

本章では、演奏者の個性を模倣した仮想演奏者とのジャムセッションを実現する上での問題点と解決法について述べる。実在する演奏者の個性が模倣可能になれば、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を模倣した個性をもつ仮想演奏者といつでもジャムセッ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。

2.1 はじめに

演奏者の個性には様々な側面があるが、たとえばジャムセッションでは、相手の演奏に対してどのような演奏で反応するかによって演奏者の個性が表れる。人間の演奏者は自分がソロを弾いているときに、相手が同時にソロを弾き出したら、次の小節から、ソロをあきらめたり、逆にもっと激しいソロを弾いたりする。そのような個性をここでは、演奏者の振る舞いと呼ぶ。演奏者の振る舞いを実在する人間の演奏者から獲得することができれば、仮想演奏者はあたかもその演奏者のように、相手の演奏を聴きそれに対してソロや伴奏でインタラクションすることが可能となる。

また、ジャムセッションではグルーブ感と呼ばれるような演奏の心地よいゆらぎや、フレーズや音自体にも個性が表れる。演奏者が実際にどのようなフレーズを弾くかは、演奏者固有のフレーズによって決まる。すべての演奏者は、手癖と呼ばれるような自分の得意とするフレーズをいくつも持っている。そのようなフレーズを獲得することができれば、仮想演奏者は模倣したい演奏者のフレーズで演奏することができる。

実際に演奏を出力する際には、演奏の発音時刻にゆらぎが生じる。このようなゆらぎにも演奏者の個性が反映されている。発音時刻ゆらぎの個性を獲得することができれば、あたかも模倣したい演奏者のように発音時刻にゆらぎのある演奏を生成することができる。

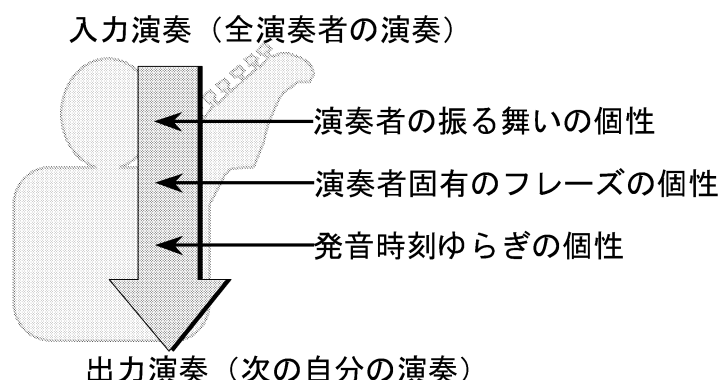


図 2.1: 演奏者の 3 種類の個性

以上述べてきた，演奏者の 3 つの個性は図 2.1 のような関係にあると考えられる．演奏者の振る舞いの個性は，一番入力に近いところにある個性であり，以下，演奏者固有のフレーズの個性，発音時刻ゆらぎの順に，出力に近くなる．本研究では，模倣したい演奏者の演奏記録から，その演奏者の個性として，上記の 3 種類の個性をそれぞれ獲得し計算機上の仮想演奏者に反映させる．

2.2 システムの全体像

本ジャムセッションシステムは，演奏者の個性を収めた 3 種類の個性データベースと，その個性を用いて動作する仮想演奏者からなる（図 2.2）．本システムの特徴は，模倣したい演奏者のジャムセッションの演奏記録さえあれば，その演奏者を模倣した仮想演奏者を生成するための個性データの獲得が可能な点である．したがって，個性データベースには，システムを使ってジャムセッションした演奏者の個性データを次々と追加していくことができる．

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 [後藤 99] の全プレー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だけでなく，人間の演奏者に入れ替わってその個性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 2.3）．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

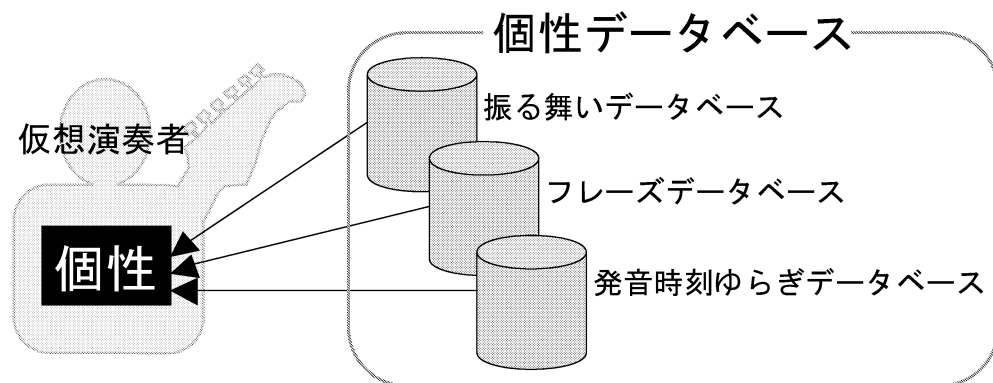


図 2.2: システムの全体像

手が人間であるか仮想演奏者であるかは問わない。

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で一定、コード進行は固定とする。システムの入出力にはMIDI¹を用いる。

2.3 演奏者の振る舞いの模倣

本研究では、全演奏者の過去の演奏から、次に自分がどのような演奏をするかを決める過程を「演奏者の振る舞い」と定義する（図 2.4）。この振る舞いを、実在する人間の演奏者から獲得することができれば、仮想演奏者は、あたかもその演奏者のように、相手の演奏を聴きソロや伴奏でインタラクションすることが可能となる。その実現には以下の課題がある。

[実現上の課題とその解決法]

² 入力の抽象化

システムが適切な出力を出すためには、セッション時のMIDI入力データと同じような入力が、学習時のMIDIデータに含まれていることが望ましい。しか

¹人間の演奏者からのMIDI入力には、MIDIギター（YAMAHA G50, G1D）、仮想演奏者の演奏出力にはMIDI音源（Roland SC-88 Pro）を用いる。



図 2.3: 人間の演奏者と仮想演奏者とは置換可能なセッションのモデル

し、実際のセッションの演奏では、MIDI データのレベルで同じ演奏が再現される確率は極めて低く、様々な入力に対応するためには、膨大な量の MIDI 入出力データの組が必要となり、入出力関係を学習することが困難である。本研究では、この問題の解決法として入力を一段抽象化した、印象空間を導入する（図 2.5）。

2 出力の抽象化

実際のセッションの演奏では、同じような印象を与える演奏に対して、演奏者が MIDI データのレベルで全く同じ演奏で反応する可能性は非常に低く、複数の出力が考えられるため、入出力関係を学習することが困難である。そこで本研究では、この問題の解決法として出力を一段抽象化した意図空間を導入する²。

以下、その具体的な解決法について述べる。

²文献 [NK89, NW92] では、入出力演奏から輪郭やリズムなどの物理的な特徴を抽出し、その入出力関係を 3 つのニューラルネットワークで学習している。学習サンプル数が少ない場合には、そのように物理的な特徴の入出力関係を学習することも可能である。しかし、本研究では、ジャムセッションの演奏記録を有効に使うことによって、より精密な演奏者モデルを構築することを目指しているため、そのような手法は適していない。

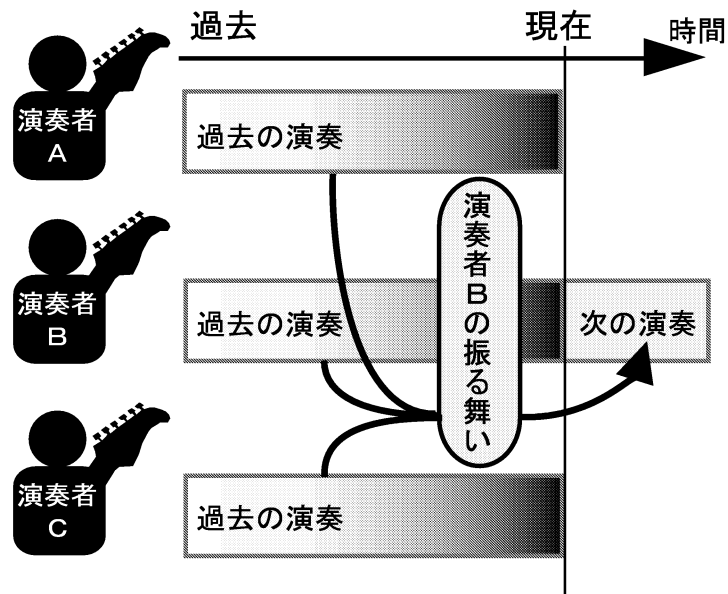


図 2.4: 演奏者の振る舞い

2.3.1 印象空間の導入

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人の演奏者の演奏から受ける印象は、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。印象空間を導入することによって、MIDI データ上では、異なる演奏であっても、それが同じ印象を与えるのであれば、同じ印象ベクトルとして表される。同じ印象を与える演奏は再現され易く、入出力関係の学習が可能となる。

2.3.2 意図空間の導入

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、被験者実験の結果求めた主観的類似度（被験者の主観に基づく類似度）に基づき配置した空間で、どのような演奏を生成したいかを表す。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

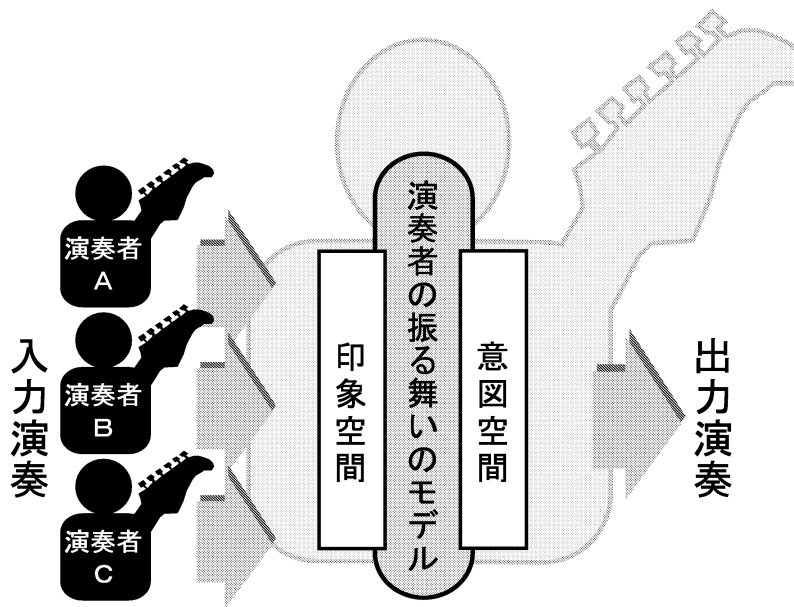


図 2.5: 演奏者の振る舞いのモデル

呼ぶ。意図空間を導入することによって、演奏者の振る舞いのモデルは、印象空間から意図空間への関数として学習することが可能となる³。

2.3.3 振る舞いのモデルの獲得

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。その具体的な手法については、3章で述べる。

2.4 演奏者固有のフレーズの模倣

演奏者の演奏記録から、フレーズを切り出し、再利用することができれば、仮想演奏者は演奏者固有のフレーズを模倣することが可能となる。本研究では、ジャム

³印象空間と意図空間は共に演奏を抽象化した空間で、両者が同じ構造の場合でも、原理的には振る舞いのモデルの学習が可能である。ただし、できる限り人間の演奏聴取過程、演奏生成過程を模したものであることが望ましいため、異なる構造の空間を採用している。演奏者は相手の演奏から受ける印象に基づき、自分の演奏意図を決定し、その意図に対応した演奏パターンを次々と出力していると、筆者は考えている。

セッションの演奏記録からフレーズデータベースを自動生成し、データベースに蓄積されたフレーズを再利用して仮想演奏者の演奏を生成する。フレーズデータベースを自動で生成するためには、以下の課題がある。

[実現上の課題とその解決法]

2 発音時刻ゆらぎの除去

ゆらぎを除去せずにフレーズの切り出しを行うと、生成時にフレーズを接続する際に、つなぎ目が不自然となる部分が生じてしまう。たとえば、はしったりリズムのフレーズと、もたったりリズムのフレーズが接続された部分では、拍の間隔が不規則となり聞き苦しい演奏となる。本研究では、フレーズ切り出しの前処理として、発音時刻のゆらぎ・ずれを除去するクオンタイズ (quantize) を行う。

2 フレーズの切り出し

フレーズ切り出しの際、小節線の位置などで機械的に演奏記録を分割すると音楽の階層構造を壊してしまう場合がある。そして、そのようなフレーズの断片を再利用することは困難である。本研究では、音楽の階層的な構造を考慮した、フレーズの切り出しを行う。

2 フレーズの意図空間への配置

切り出したフレーズを再利用して仮想演奏者が演奏を生成するためには、仮想演奏者がどのような意図のときにそのフレーズを使用すれば良いか分からなければならない。本研究では、各フレーズに対応する意図ベクトルを求め、意図空間上に配置する。

以下、その具体的な解決法について述べる。

2.4.1 発音時刻のクオンタイズ

発音時刻を推定しクオンタイズすることで、発音時刻のゆらぎを除去する。ジャムセッションのように、8分3連音符や16分音符が頻繁に入れ替わるような演奏は、従来の閾値処理による機械的なクオンタイズで正しくクオンタイズすることは困難である。そこで本研究では、確率モデルを用いて、ジャムセッション等で演奏者が弾

いた演奏の発音時刻から，元々演奏者が弾こうとした正規化された楽譜上の発音時刻を推定する手法を提案する．この手法の特徴は，モデルパラメータをジャムセッションの演奏記録から統計的に学習することにより，各演奏に適した確率モデルを使ってクオンタイズできる点である．その具体的な手法については，4章4.2節で述べる．

2.4.2 フレーズの分割

音楽の階層的な構造を考慮しながらフレーズ分割を行う．そのようなことをするための音楽理論としては，GTTM(Generative Theory of Tonal Music)[LJ83]のグルーピング構造分析がある．しかし，GTTMのグルーピング構造分析の対象は単旋律で，和声に関する分析が欠如しているため，ジャムセッションの演奏のフレーズ分割に直接用いることはできない．また，ルールが定量的に定義されておらず，複数のルール間の順位が不明確で競合が起きるなど，理論を計算機上へ実装することは困難だと考えられている．そこで本研究では，GTTMによるグルーピング結果と近い結果を出すボロノイ線図を用いたグルーピング手法を提案し，フレーズ分割を行う．その具体的な手法については，4章4.3節で述べる．

2.4.3 フレーズの再利用

切り出した多くのフレーズを意図空間上に配置するため，多次元尺度法の1つであるMDA-ORを用いて，フレーズの配置を行う．そして，新たに追加したフレーズの意図空間上の位置は，局所自己相関関数によって得られたフレーズの物理的特徴から決定する．仮想演奏者のそのときの意図に対応するフレーズを次々と接続することで，出力演奏を生成する．その具体的な手法については，4章4.4節で述べる．

2.5 発音時刻のゆらぎの模倣

演奏者の発音時刻の特徴・癖がわかれば，ゆらぎの無い演奏にゆらぎを加え，あたかも実在する演奏者のように，ゆらぎのある演奏を生成することができる．そのためには，以下の課題がある．

[実現上の課題とその解決法]

2 発音時刻ゆらぎの個性をどのように獲得するのか

ゆらぎの無い演奏にゆらぎを付加するためには、まず演奏者の発音時刻の特徴や癖をなんらかの形で獲得しなければならない。ここでは、クオンタイズで演奏記録から学習した確率モデルに演奏者の発音時刻ゆらぎの個性が反映されていると考え、それを使って発音時刻にゆらぎを付加する。

2 発音時刻ゆらぎの個性をどのように反映させるのか

発音時刻ゆらぎの模倣を実現するためには、確率モデルに基づいたゆらぎ作成し、それをういて発音時刻にゆらぎを与えなければならない。ここでは、逆関数法を用いてそれを実現する。

以下、その具体的な解決法について述べる。

2.5.1 ゆらぎの個性の獲得

クオンタイズで演奏記録から学習した確率モデルを発音時刻ゆらぎの個性として用いる。ただし、クオンタイズで獲得した確率モデルには、学習データの不足により分布にばらつきがあるとともに、演奏者のミスにより、大きく発音時刻がずれている音も含まれている。したがって、その分布を用いてゆらぎを付加すると、出力にばらつきが出てしまう。そこでここでは、正規分布により近似した分布を使って、ゆらぎの付加を行う。その具体的手法については、5章5.3節で述べる。

2.5.2 逆関数法によるゆらぎの付加

逆関数法により、確率分布に基づく乱数を発生させ、演奏に発音時刻ゆらぎを付加する。このとき、和音の発音順序が入れ替わったりしないように、和音の場合には、分布を和音の構成音の個数で等分割した分布を用いて、ゆらぎを付加する。その具体的手法については、5章5.4節で述べる。

2.6 まとめ

本章では、ジャムセッションにおける演奏者の個性として、演奏者の振る舞いの個性、演奏者固有のフレーズの個性、発音時刻ゆらぎの個性の3種類を取り上げ、それを模倣する上での問題点と解決法およびシステムの全体像について述べた。システムは、ジャムセッションの演奏記録から獲得した演奏者の個性を収めた、3種類の個性データベースと、その個性を用いて動作する仮想演奏者からなる。以下、3章では演奏者の振る舞いの模倣、4章では演奏者固有のフレーズの模倣、5章では発音時刻ゆらぎの模倣について具体的に説明する。

第3章 演奏者の振る舞いの模倣

本章では、演奏者の個性として、演奏者の振る舞いを演奏記録からの統計的学習により獲得し、実在する人間の演奏者の振る舞いを模倣した仮想演奏者を実現する手法について述べる。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)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、獲得することができる。

3.1 はじめに

演奏者の振る舞いの学習は、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定式化できる。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、膨大な量のデータセットが必要となり、限られた長さの演奏記録からモデルを獲得することが困難である。そこで、本研究では、入力と出力の間に、印象空間と意図空間という2つの主観空間(主観を表す空間)を導入し、印象ベクトルから意図ベクトルへの写像を演奏者の振る舞いのモデルとして獲得する。そして、ギタートリオの演奏を対象とし、振る舞い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 を用いて実装する。以下3.2節では、振る舞いのモデルを学習する手順について説明し、3.3節では、仮想演奏者の構成について述べる。そして、3.4節では、実験により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。

3.2 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3.1）．

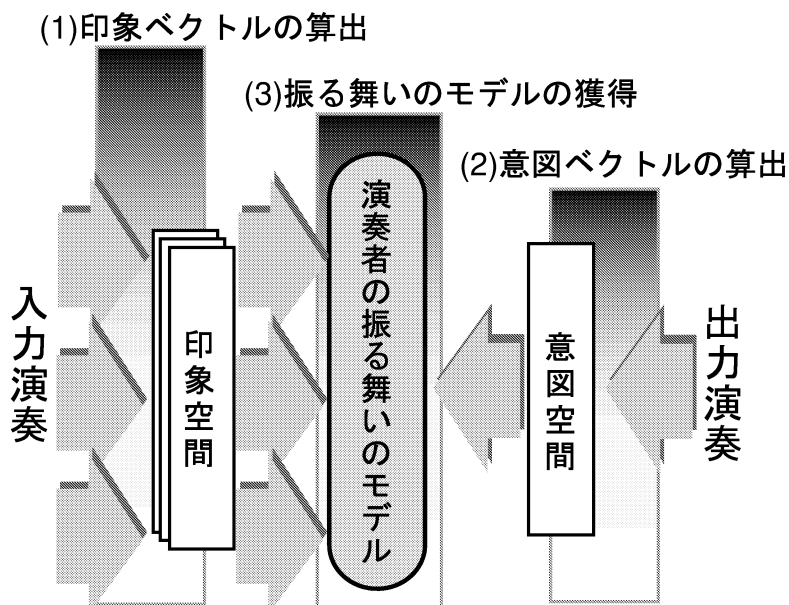


図 3.1: 振る舞いのモデルの学習の手順

3.2.1 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．

3.2.2 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，どのような演奏を生成しようとする意図であったか」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．

3.2.3 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．

3.3 仮想演奏者の構成

仮想演奏者は，1) 印象算出部，2) 意図決定部，3) 演奏生成部，の3つからなる(図3.2)．そして仮想演奏者は，振る舞いのモデルを学習するとき(学習モード)とそのモデルを用いて実際にセッションするとき(演奏モード)では，それぞれ異なった処理をする．学習モードは，非リアルタイムで実行され，演奏記録から得られる印象ベクトルから意図ベクトルへの関数はRBFネットワークを用いて学習する．一方，演奏モードは，リアルタイムで実行され，学習によって得られたネットワークを用いて各拍ごとに仮想演奏者の意図ベクトルを算出し，出力演奏を生成する．

3.3.1 印象算出部

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を正準相関分析することにより求める．正準相関

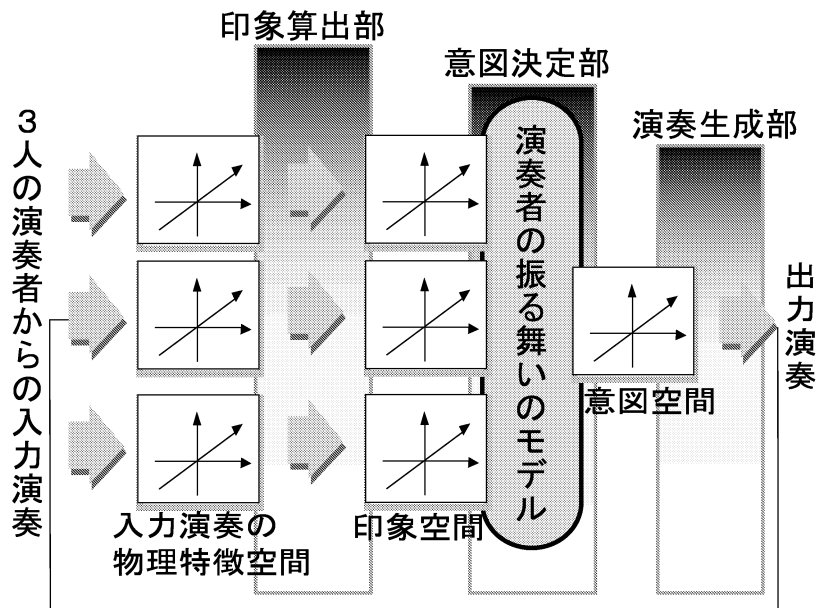


図 3.2: 仮想演奏者の構成

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を効率良く学習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ていき、正準得点の大きさによって物理特徴量を取捨選択することにより低次元化

した。

印象ベクトルの獲得

2 心理実験

被験者（音楽経験者）に演奏を聴かせながら，1小節の1/12の時間分解能で，演奏を表す10語¹（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を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を右から左に進行していく縦棒に沿って，その時刻における評定を次々と入力する（図3.3）。具体的には，印象語が適していると思った場合には，縦棒の一番上の位置をクリックし，適していないと思った場合には，縦棒の一番下の位置をクリックする。一度に画面上に表示できるのは，1コーラス分の評定である。演奏が次のコーラスに移ると，画面は次の表示に切り替わる。

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がちなので，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

¹これらは，演奏者向けの音楽雑誌で使用頻度の高かった，演奏を表す語である。頻度を求めるために使用した音楽雑誌は，ギターマガジン（Guitar magazine）[リッ98]，プレイヤー（Player）[プレ98]，JAZZ 第9巻ジャンゴ・ラインハルト[今田95]，パーン（BURRN!）[パー97]である。



図 3.3: 印象語の評価に用いたアプリケーション

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（付録A参照）。

入力演奏の物理特徴ベクトルの獲得

2 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、以下のような処理を行った。まず、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小節ごとに分割する。そして、小節内のすべての音に対し、1/48小節の分解能でベロシティー、音高、ピッチベンドの値を求め、それぞれ音量フィルタ（3種類）、音域フィルタ（3種類）、Noteフィルタ（512種類）にかける。各フィルタはそれぞれ、ベロシティーや音域、コードの主音からの音程が、ある範囲にあるときのみ1を出力する関数で、様々な物理特徴の分離、抽出を可能としている。このようなフィルタの設計は、どのような物理特徴量が演奏から受ける印象に影響を与えているかを調べた予備実験の結果に基づいている（付録B参照）。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（12小節×3（ベロシティー、音高、ピッチベンド）×3音量フィルタ×3音域フィ

ルタ × 512Note フィルタ × 4 (平均値 , 最大値 , 最小値 , 傾き) 生成された (図 3.4) .

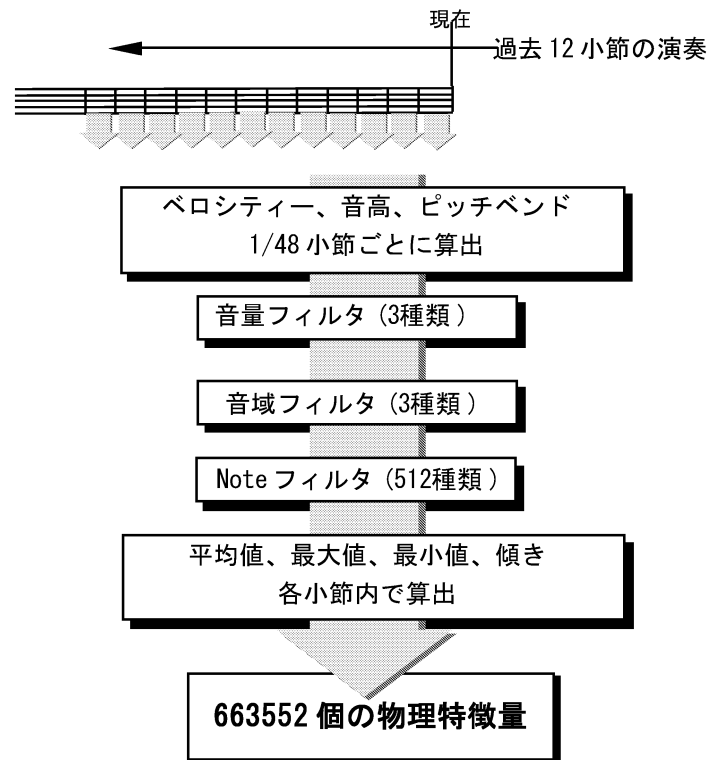


図 3.4: 様々な物理特徴量の抽出

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 (付録 C 参照) .

印象ベクトルの算出

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 3.5 は、存在感について、被験者の評定値とシステムの算出値とを、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小節に12回印象ベクトルを算出している。

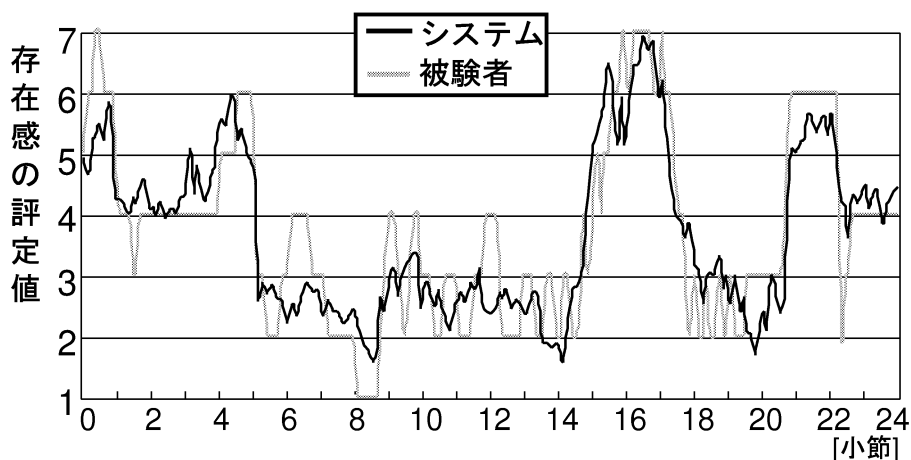


図 3.5: 被験者とシステムが示した存在感の値の比較

3.3.2 意図決定部

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。この部分では、過去の印象ベクトルを入力し意図ベクトルを出力するようなネットワークを構成する。

このようなことを行う学習アルゴリズムとして、本研究ではRBF (Radial Basis Function) ネットワークを採用した。RBF 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の印象ベクトル x (L 次元) を入力

し、3次元の意図ベクトル $y_1; y_2; y_3$ (3次元) を出力するような RBF ネットワークを構成する (図 3.6) . 意図ベクトルの次元数は、多次元尺度法により 3次元としており、具体的には、3.3.3 で述べる .

学習用データは、12 コーラスの演奏記録に含まれる、過去の印象ベクトル x とそれに対応する意図ベクトル $y_1; y_2; y_3$ との対、 N 個 ($11(=12-1)$ コーラス \times 12 小節 \times 4 拍 = 528 個) である . したがって、入出力データの組は $(x^{(k)}; y_j^{(k)}) (k = 1; 2; \dots; N; j = 1; 2; 3)$ となる .

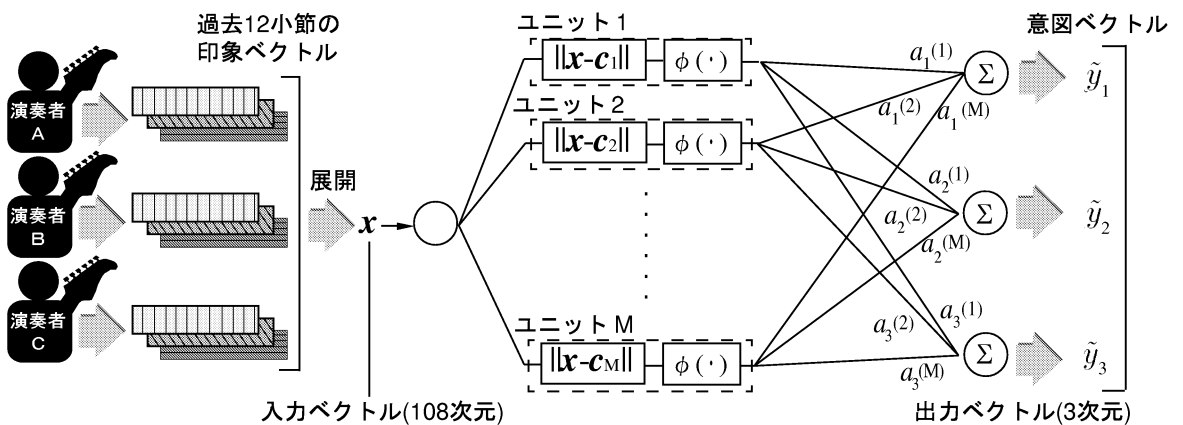


図 3.6: 振る舞いのモデルを学習する RBF ネットワーク

ネットワークは M 個 ($M \geq N$) のユニットを持ち、各ユニットの中心 $c_i (i = 1; 2; \dots; M)$ は L 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 $x^{(k)}$ とユニットの中心 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 $A(r)$ を出力する (式 (3.1)) .

$$A(r) = \exp \left\{ -\frac{r^2}{2\sigma^2} \right\} \quad \sigma \neq 0, r \geq 0 \tag{3.1}$$

$$r = \sqrt{\sum_{j=1}^L (x_j^{(k)} - c_{ij})^2}$$

ネットワークの出力 $y_j^{(k)}$ はそれに重み $a_j(i)$ を掛け加え合わせたものである (式 (3.2)) .

$$y_j^{(k)} = \sum_{i=1}^M a_j(i) A \left(\sqrt{\sum_{j=1}^L (x_j^{(k)} - c_{ij})^2} \right) \tag{3.2}$$

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。

$$y_j = P(x)a_j + e_j \quad (3.3)$$

ただし、

$$P(x) = \begin{matrix} & \begin{matrix} 2 & & & 3 \\ p_1^{(1)} & p_2^{(1)} & \dots & p_M^{(1)} \\ p_1^{(2)} & p_2^{(2)} & \dots & p_M^{(2)} \\ \vdots & \vdots & & \vdots \\ p_1^{(N)} & p_2^{(N)} & \dots & p_M^{(N)} \end{matrix} & \end{matrix}$$

$$p_i^{(k)} = \sum_j A_{ij} x_j^{(k)} + c_{ij} \quad (k = 1; 2; \dots; N; i = 1; 2; \dots; M)$$

$$y_j = \begin{matrix} \begin{matrix} 2 & & & 3 \\ y_j^{(1)} \\ y_j^{(2)} \\ \vdots \\ y_j^{(N)} \end{matrix} ; a_j = \begin{matrix} 2 & & & 3 \\ a_j^{(1)} \\ a_j^{(2)} \\ \vdots \\ a_j^{(M)} \end{matrix} ; e_j = \begin{matrix} 2 & & & 3 \\ e_j^{(1)} \\ e_j^{(2)} \\ \vdots \\ e_j^{(N)} \end{matrix}$$

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 [CCG91] (付録 D 参照)。

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 (式 3.4)。

$$P = WB \quad (3.4)$$

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．

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 (3.2) より 2 乗誤差は以下で与えられる．

$$\begin{aligned} \|e_j\|^2 &= \|y_j - P(x)a_j\|^2 \\ &= (y_j - P(x)a_j)^T (y_j - P(x)a_j) \end{aligned} \quad (3.5)$$

これを a_j の各成分で偏微分すると，

$$\frac{\partial \|e_j\|^2}{\partial a_j} = -2 P(x)^T y_j + 2 P(x)^T P(x) a_j = 0 \quad (3.6)$$

したがって， a_j の推定値は式 (3.7) で与えられる．

$$a_j = (P(x)^T P(x))^{-1} P(x)^T y_j \quad (3.7)$$

L の設定，M の最適化

RBF ネットワークの入力次元数 L を設定し，ユニット数 M ，式 (3.1) の σ の最適な値を求める．RBF ネットワークの汎化能力は，ユニット数 M および σ によって大きく変化する．また，印象ベクトルの次元数 L も適切な値に設定しなければ，モデル化誤差が大きくなる． L の次元数は，大きいほど過去の演奏履歴をより反映した演奏ができるため望ましいが，144 次元（3 人 × 16 小節 × 3 次元）以上ではモデル化誤差が大きかったため，108 次元（3 人 × 12 小節 × 3 次元）とした． L を 108 次元とし， M 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，モデル化誤差が最も少なくなるパラメータを求めたところ， $M = 46$ ， $\sigma = 23.5$ が最適な値として求まった．図 3.7 は $L=108$ のときに M および σ を変化させたときのモデル化誤差の値を示したものである．他のサンプルデータについても，同様のパラメータの値で重みを計算できたため，演奏者の振る舞いの学習を行う際には，常にこの値を用いることにした．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．

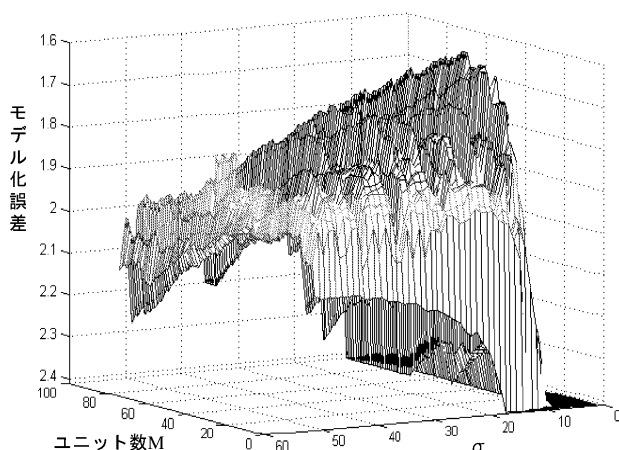


図 3.7: ユニット数 M と σ を変化させたときのモデル化誤差の値

3.3.3 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。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。

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手作業で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 (図 3.8)。

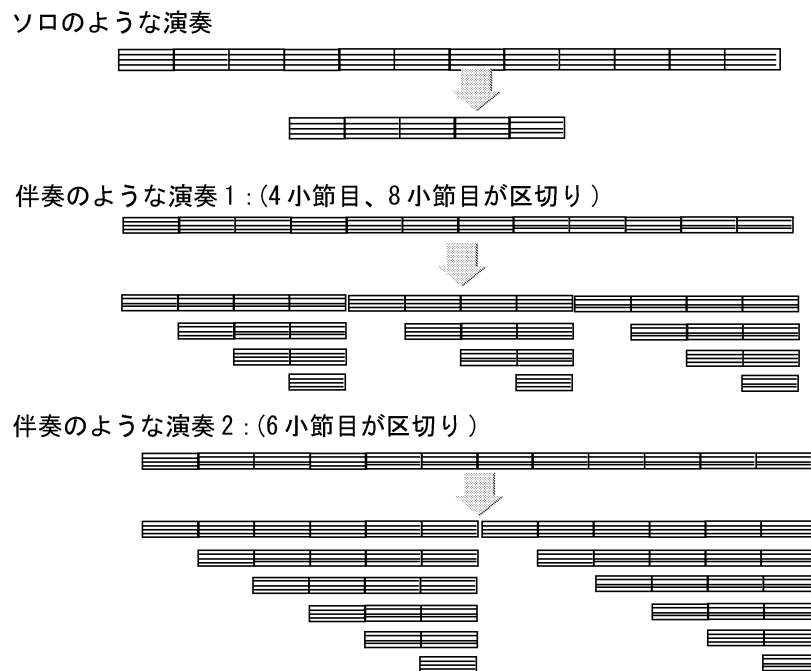


図 3.8: 演奏パターンの生成

一対比較実験

被験者（3.1.1 節の被験者と同一人物）に，77 個²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 : 「似ていない」，2 : 「少し似ていない」，3 : 「少し似ている」，4 : 「似ている」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．

多次元尺度法

Kruskal の多次元尺度法 [KW78] 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，演奏パターン j と演奏パターン k の類似度を r_{jk} ，多次元空間での距離を d_{jk} としたとき，類似度の低い演奏フレーズほど

²77 個のパターンの内訳は，1 コーラスのすべての小節から始めるソロのようなパターンを 5 種類ずつ計 60 個（12 小節 × 5 種類），A，E，D，それぞれのコードでの伴奏のようなパターンを 5 種類ずつ計 15 個（3 コード × 5 種類），2 種類のターンアラウンドである．

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手法である（式（3.8））。

$$\pm_{jk} > \pm_{lm} \quad \text{ならば} \quad d_{jk} < d_{lm} \quad (3.8)$$

このとき，式（3.8）が成立しない度合いは，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．ストレス値は，次式で表される．

$$S = \frac{2 \sum_j \sum_k (d_{jk} - \pm_{jk})^2}{4 \sum_j \sum_k d_{jk}^2} \quad (3.9)$$

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．

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得られる．逆写像は，陽に求められ，演奏者の振る舞いのモデルを学習するとき（学習モード）と，新たな演奏パターンをデータベース上に加えるときに用いる．

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により生成される．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．データベースにある演奏パターンの半数はソロ演奏であり，仮想演奏者はどの小節からでも自分の意図に近いソロ演奏を開始することができる．

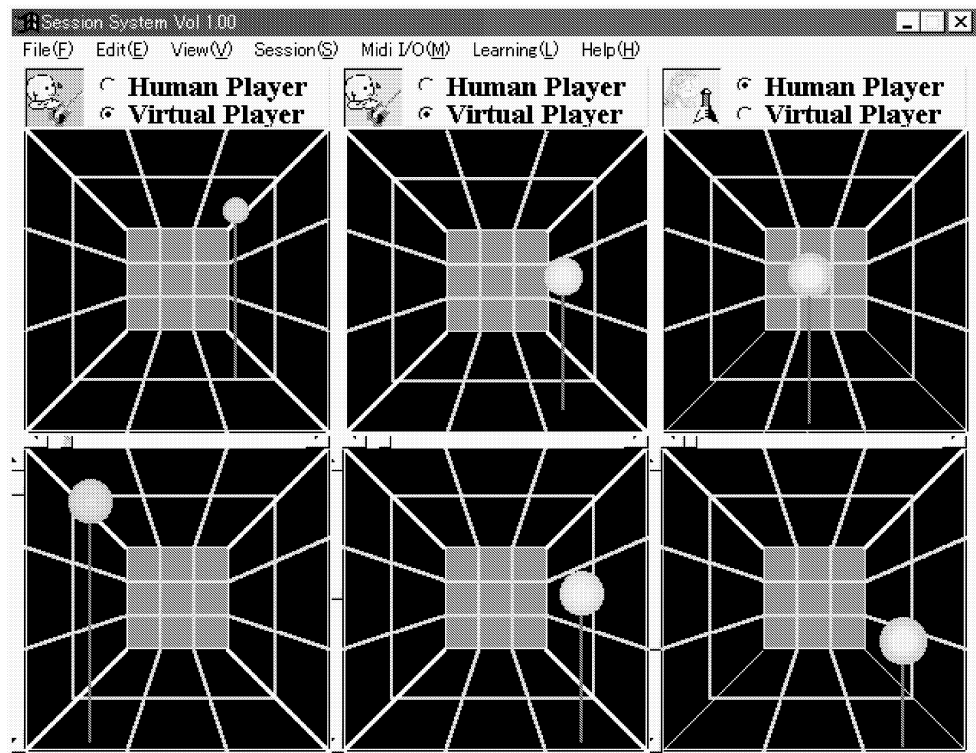


図 3.9: システムの出力画面

3.4 実験結果

システムをパソコン（Pentium 650MHz，Windows98）上に実装した．図 3.9 は，システムの出力画面である．画面は，横に 3 つのパネルが並んでおり，各パネルが各演奏者に対応している．パネルの一番上にあるラジオボタン（選択ボタン）は，そのパネルを仮想演奏者が使用するか，人間の演奏者が使用するかを切り替えるためである．そして，ラジオボタンの下には，2 つの立方体が描かれている．中央の薄い色の正方形が立方体の一番奥にある面で，それを取り囲む 4 つの台形が立方体の上下左右の面である．上の立方体は印象空間を表し，その中にある球の位置が，印象ベクトルを表している．下の立方体は，意図空間を表し，その中にある球の位置が，意図ベクトルを表している．

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

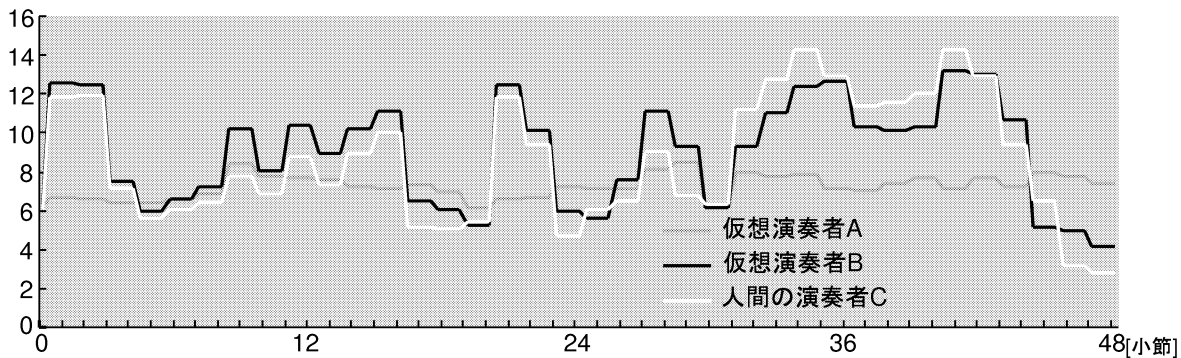


図 3.10: 3 人の演奏者の意図ベクトルの変化 (3 軸の意図ベクトルのうち 1 成分のみを表示)

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。仮想演奏者 A, B は、人間 3 人で行った同じセッションの記録から学習したものである。

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 3.10 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。

さらに、5 人のギタリストにシステムとセッションした感想を求めたところ、仮想演奏者は 1 人 1 人それぞれ個性的な振る舞いをするとの意見が多かった。特に、自分自身を模倣した仮想演奏者とジャムセッションした演奏者は、仮想演奏者が自分の真似をしているようで演奏しにくいと感じた者が多かった。これは、仮想演奏者の RBF ネットワークが、人間の演奏者の演奏意図を予測できていたためだと考えられる。

3.5 まとめ

本章では，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を実現する手法について述べた．本システムでは，全演奏者の演奏履歴と，模倣しようとしてい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を獲得することを可能にした．実験の結果，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が確認できた．

第4章 演奏者固有のフレーズの模倣

本章では、演奏者の個性として、演奏者固有のフレーズを、模倣したい演奏者の演奏記録から切り出し、再利用することによって、演奏者固有のフレーズの模倣を実現するための手法について述べる。Sloboda は文献 [Slo87] で、「即興演奏では、限られた時間内にある結果を生み出す必要があり、数多くの手続きや方法が'ただちに利用できる'状態にあることから生まれてくる」と述べた。演奏者が、多くのフレーズの中から適切なものを選択し、それを接続していくことにより演奏が生成可能であるという本研究の考え方の根拠は、即興演奏についての Sloboda のこの仮説にある。十分な数のフレーズをあらかじめ仮想演奏者に用意しておけば、相手の演奏に対してただちに適切なフレーズで反応することが可能となる。

4.1 はじめに

3章で述べたジャムセッションシステムは、事前に人間の演奏者から学習したモデルに基づき演奏意図を決定し、フレーズデータベースの中からそのときの意図に対応したフレーズ(演奏パターン)¹を次々と接続することで演奏を生成していた。フレーズデータベースには、1小節から8小節の長さの演奏が収められているが、それらは手作業で作成していたため、仮想演奏者は演奏者固有のフレーズまでは模倣できなかった。本章では、模倣したい演奏者の即興演奏からフレーズを切り出して、データベースを自動作成することによりそのような模倣を可能とする。フレーズデータベースを作成するためには、まず、ゆらぎを含む即興演奏から確率モデルに基づくクオンタイズ手法(4.2節)を用いてゆらぎを除去する必要がある。次にポロノイ線図に基づくフレーズ分割手法(4.3節)を用いて、実際に演奏記録からフレーズを切り出す。最後に、フレーズの再利用法(4.4節)を用いて、切り出したフレーズを

¹3章で述べた演奏パターンという語と、本章でのフレーズという語は同じ意味である。本章の内容にはフレーズという語のほうがより適切だと考える。

仮想演奏者の意図空間上に配置する。

4.2 発音時刻のクオンタイズ

本節では、即興演奏のフレーズ分割を自動化するため、テンポ一定の伴奏演奏（たとえばドラムスによる伴奏）に合わせて弾かれた演奏の発音時刻を、楽譜上の正規化された位置へとクオンタイズする手法について述べる。つまりここで扱うのは、ゆらぎを含むような即興演奏からゆらぎを取り除く問題である。ただし、即興演奏の場合には楽譜を演奏するわけではないため、楽譜上の正規化された位置は、元々演奏者が弾こうとした小節・拍内の量子化された位置を意味する。

市販のシーケンスソフトウェアに搭載されているクオンタイズは、ユーザがグリッド間隔（分解能）を指定し（たとえば、8分3連音符、16分音符等の一定間隔のグリッドを指定し）、発音時刻を最も近いグリッドへ整列させるものである。したがって、拍構造が既知で固定されている演奏に対しては有効だが、ジャムセッションのように、8分3連音符や16分音符が頻繁に入れ替わるような演奏を正しくクオンタイズするためには、演奏の部位ごとに人間が手作業で指定する必要があり、我々のフレーズ分割を自動化する目的には使えない。

従来、コネクショニストモデルによるクオンタイズの研究 [DH89]、音響信号に対するクオンタイズの研究 [片寄 90]、ビート・リズムの認識に関するビートトラッキングの研究 [後藤 98a, GM95, 後藤 98b, DMR87, AD90] 等がなされてきた。文献 [DH89] の研究は、隣接する発音時刻間隔（あるいはその和）の比が整数比になれば安定するエネルギー関数を定義し、繰り返し計算で発音時刻のずれを修正していたが、エネルギー関数が固定されていたため、多様な演奏へ対処することは困難であった。また、文献 [片寄 90] の研究は自動採譜のための拍の推定のためのものであり、文献 [後藤 98a, GM95, 後藤 98b, DMR87, AD90] の研究は拍位置を予測することに主眼が置かれていた。いずれも、問題設定が本研究（テンポ一定の伴奏に合わせて弾かれた演奏のクオンタイズ²）とは異なっており、直接それらの手法は適用できない。

一方、文献 [大槻 02] では、本研究に近い問題設定が扱われている。この研究は、連続音声認識の方法論が、音符推定・テンポ推定・拍子推定にも有効であることを

²このような問題設定でも、単純な閾値処理は有効でなく（ゆらぎによるずれを誤ってクオンタイズする等）、難しい課題である。

示した点が優れており，音符長の遷移と変動を隠れマルコフモデル（HMM）でモデル化することで，市販のシーケンスソフトウェアのクオンタイズを凌ぐ性能を得ることに成功している．しかし，単音の音符列に対する検討しか行っていなかった．

本研究では，発音時刻に焦点を絞った発音時刻の遷移とゆらぎのモデルを新たに提案し，和音を含む演奏に対しての評価を行う．さらに，教師データが与えられない条件で，確率モデルのモデルパラメータを推定する手法を提案し，教師データが与えられない場合でもクオンタイズできることを示す．本手法の特徴は，モデルパラメータを演奏記録から統計的に学習することによって，各演奏に適した確率モデルを使ってクオンタイズすることができる点である．以下，4.2.1ではクオンタイズの問題を逆問題としてとらえる立場から，隠れマルコフモデルを用いた問題の定式化について述べ，4.2.2では実際のクオンタイズの手法について述べる．4.2.3では，モデルパラメータの教師つき・教師なし学習法について述べ，4.2.4では，実験により提案モデルがシーケンスソフトウェアのクオンタイズを越える性能を持つことを示す．

4.2.1 学習に基づくクオンタイズ

人間の演奏者が同じ演奏をくりかえし弾いた場合でも，MIDIのレベルでまったく同じ演奏となることは稀であり，演奏動作の微妙な差や，演奏の表情づけの差などにより発音時刻がゆらぐ．これを，元々演奏者が弾こうとした発音時刻（小節・拍内の正規化された位置）系列から，ゆらぎのある実際に演奏された発音時刻系列を求める順問題とすると，演奏された発音時刻系列から演奏者が元々弾こうとした正規化された発音時刻系列を求めるクオンタイズは，その逆問題³となる（図4.1）．ここでは，前者のゆらぎの生じる過程を「発音時刻ゆらぎの順モデル」としてモデル化する．そして，この順モデルから求めた「発音時刻ゆらぎの逆モデル」を用いて逆問題を解き，クオンタイズを実現する．

³このように，実際の演奏から，その拍構造，楽譜構造を推定する問題を逆問題として捉える考え方は，文献 [後藤 98a, GM95, 後藤 98b] で紹介された．文献 [大槻 02] でも，同様に逆問題と捉えて定式化しており，本研究でもそれを参考にしている．

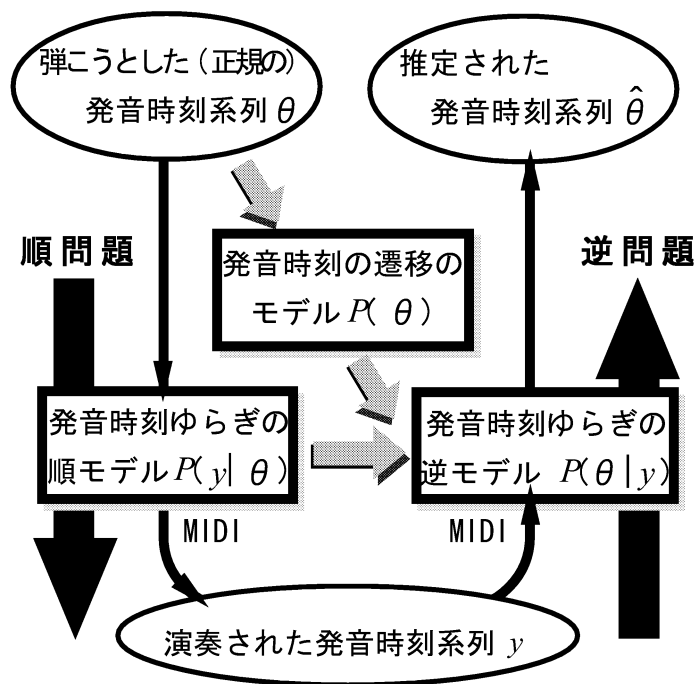


図 4.1: クオンタイズにおける順モデルと逆モデル

発音時刻の遷移とゆらぎのモデル

発音時刻ゆらぎの順モデルは、元々弾こうとした正規化された発音時刻系列を μ 、演奏された発音時刻系列を y とした時、 $P(y|\mu)$ で表される。このとき、逆モデルは、Bayes の定理より式 (4.1) で表される。

$$P(\mu|y) = \frac{P(y|\mu)P(\mu)}{P(y)} \quad (4.1)$$

$P(\mu)$ は、 μ に対する事前分布であり、演奏者がどのような発音時刻系列を弾きやすいかを表わす。

逆問題の解 $\hat{\mu}$ は、式 (4.1) を最大化する μ である (式 (4.2))。

$$\begin{aligned} \hat{\mu} &= \operatorname{argmax}_{\mu} P(\mu|y) \\ &= \operatorname{argmax}_{\mu} P(y|\mu)P(\mu) \end{aligned} \quad (4.2)$$

隠れマルコフモデルによる定式化

本研究では、発音時刻ゆらぎの順モデル $P(y|\mu)$ と発音時刻の遷移のモデル $P(\mu)$ を組み合わせたモデルを隠れマルコフモデル (HMM) で定式化する。隠れマルコ

フモデルはマルコフ的な隠れ状態遷移モデル $P(\mu)$ と各状態における出力確率分布 $P(y_j|\mu)$ を組み合わせたモデルである。我々が観測することができるのは、出力確率分布から得られる出力のみで、どの状態にいるかを観測することはできない。

2 演奏のモデル化

演奏をモデル化する前提条件として、以下のものを考える。

1. 何をモデル化するのか

本研究では、演奏の発音時刻すなわち音の立ち上がり時刻の分布をモデル化する。文献 [大槻 02] では、音符の長さをモデル化していたが、ギターのアルペジオや開放弦を含む演奏や、ピアノで右手と左手を同時に弾いているような演奏では、音符の長さのみで、演奏をモデル化することは困難であり、発音時刻でモデル化の方が有効であると考えられる。

発音時刻を用いて演奏をモデル化する際、発音時刻の前後関係が保存されるようにすることで、音の順序が反転することがない妥当なクオンタイズを実現できる。

2. 演奏のモデル化の単位

本研究では、演奏をモデル化する単位として1拍の長さを採用した。1拍単位でモデル化を行う理由は、市販のクオンタイズを使った場合の主要な問題点である8分3連音符と16分音符の識別が1拍の長さで可能となるためである（8分3連音符は1拍を3等分した位置、16分音符は1拍を4等分した位置に出現する。）実際に曲をクオンタイズする場合には、1拍のモデルを、次々とつなげあわせて曲全体を表現する。1拍という短い長さでモデル化することは、計算時間の節約と、学習の際により多くの演奏データを用意しやすいという利点をもつ。

3. 演奏の量子化単位

以下の2種類の異なる量子化単位を持つ時刻を定義する。

拍内のイベント時刻 k

実際に演奏された発音時刻をイベント時刻と呼ぶ。その量子化単位は、シーケンスソフトウェアで採用されることが多い1拍の1/480を単位とする。拍内のイベント時刻 k は、その拍内での時刻を表わ

し、伴奏の拍線（拍の境目）を基準に0を決め、0から479の整数値をとる。

正規化後の拍内位置 i

演奏者が元々弾こうとした正規化された発音時刻を表わす。量子化単位は、8分3連音符と16分音符の両方に対応できるように1拍の1/12を単位とし、 i は1から12の整数値をとる。

本研究で述べるクオンタイズとは、演奏を通しての観測されたイベント時刻の系列から正規化後の拍内位置の系列を求めることである。

2 1拍のHMMモデル

隠れマルコフモデルの隠れ状態を、正規化後の拍内位置に対応づけ、モデルの出力を観測されたイベント時刻、すなわち、実際の発音時刻に対応づける。ここでは、拍を12等分した位置を、拍の頭から順に状態1から状態12までに対応させる。遷移は仮想的な状態であるStartからはじまって、Endで終了する。和音など、同じ発音時刻で複数の音出力される演奏の場合は、同じ状態を自己ループする遷移となる。以下に、遷移の例をあげる。

拍内に8分3連音符が3つ並んでいる場合

Start 1 5 9 End

拍内に16分音符が4つ並んでいる場合

Start 1 4 7 10 End

拍の頭で2音同時に鳴る和音の場合

Start 1 1 End

拍内に音が無い場合

Start End

HMM（離散HMM）の各パラメータを以下のように意味づけた。

- ・隠れ状態 i ：発音時刻の正規化後の拍内位置（ i は1から12の整数）
- ・出力 k ：拍内のイベント時刻（ k は0から479の整数）
- ・状態遷移確率 a_{ij} ：拍内の i の位置で発音した後 j の位置で発音する確率。遷移は、常に拍内の前から後ろに向かって進むため、 $a_{iS} = 0$ 、 $a_{Ei} = 0$ とな

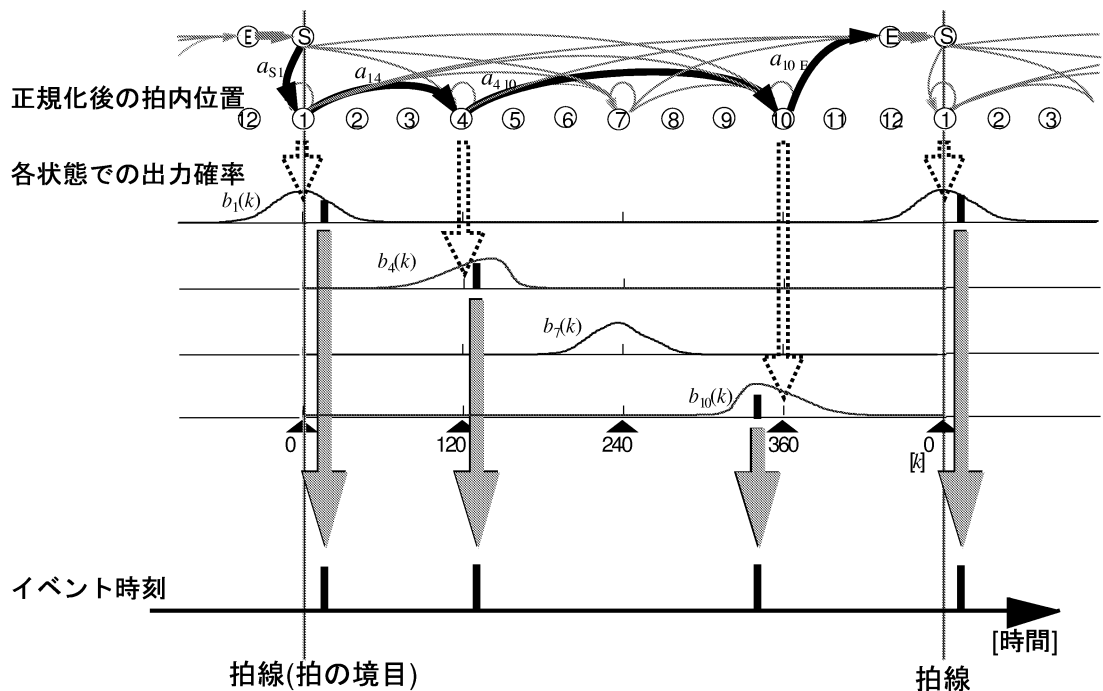


図 4.2: 1 拍の隠れマルコフモデルの概略図

り, $i > j$ のときは $a_{ij} = 0$ となる.

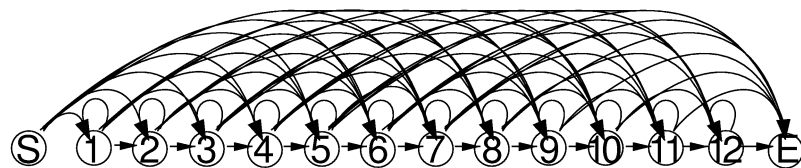
・出力確率 $b_i(k)$: 発音時刻の正規化後の拍内位置が i のときに, 拍内のイベント時刻が k となる確率. 状態 Start と End は出力を出さないため, 対応する出力分布は存在しない. 状態遷移は常に S から始まるため, HMM 初期状態分布 π は, $\pi_s = 1, \pi_i (i = 1; 2; \dots; 12; E) = 0$ となる.

図 4.2 に, 本研究における HMM の概略図を載せる. 図では, 簡単のため 12 個の状態のうち, 1, 4, 7, 10 以外の状態に関する遷移と出力確率は省略している.

2 複数の HMM によるモデル化

以上のような隠れマルコフモデルのすべての遷移を考えた場合, 図 4.3(a) のようにアークの本数が非常に多く複雑となり, 遷移確率の統計的な推定精度が低くなると考えられる. しかしここで, 演奏者が, 1 拍の内部を 16 分音符で演奏しようとする場合と, 8 分 3 連音符で演奏しようとする場合とを分けて考えると, 両者はより単純になる. つまり, 16 分で演奏しようとする場合には, 状態 1, 4, 7, 10 へ遷移する確率が高くなり, 8 分 3 連で演奏しようとする

(a)1つのHMMによるモデル



(b)4つのHMMに分岐したモデル

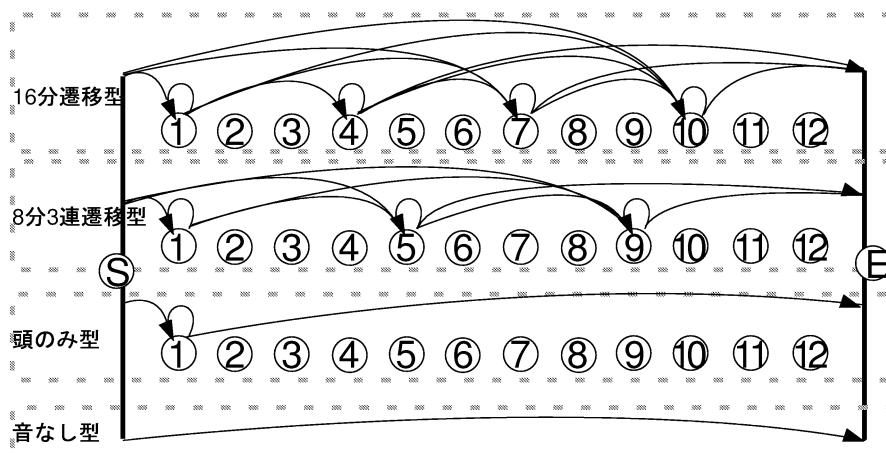


図 4.3: 1つの HMM によるモデルと 4 種類の HMM に分岐したモデル

る場合には、状態 1, 5, 9 へ遷移する確率が高くなる。したがって、1 拍を 1 つの HMM によってモデル化するよりも、複数の HMM を切り替えて用いるほうが、より詳細なモデル化ができ、クオンタイズの性能がさらに向上すると期待される。そこで、本研究では、図 4.3 のように、1 つの HMM によるモデルと共に、16 分音符で弾かれている「16 分遷移型」、8 分 3 連音符で弾かれている「8 分 3 連遷移型」、拍の頭だけ弾かれている「頭のみ型」、何も弾かれていない「音なし型」の 4 種類の HMM に分岐したモデル⁴を用い、実験によりどちらのモデルが高い性能であるかを比較する。4 つの HMM によるモデルは、1 拍の演奏が 4 種類のモデルのいずれかに当てはまるという制約を与えたものだといえる。通常、1 拍の内部に 16 分系の音符と 8 分 3 連系の音符が混在することは考えにくいので、このような制約は妥当であると考えられる。また、本研究で実験に使った演奏には、32 分音符など 8 分 3 連や 16 分よりも短

⁴このような複数の HMM によるモデル化は文献 [大槻 02] でも扱われており、有効性が確認されている。

い音価の音符は含まれていなかった。

4.2.2 最適発音時刻系列の推定

1曲を通して、最適な状態遷移系列を推定するために、各拍ごとのモデルの Start と End をつなぎ合わせた HMM を用いて、事後確率 $P(\mu_j | y)$ を最大にする状態遷移系列を探索する。そのための方法としては Viterbi アルゴリズムを用いる。

T 個の音からなる演奏を考え、演奏を通しての観測されたイベント時刻の系列を $y = (y_1; y_2; \dots; y_T)$ とする。これに対する最適状態遷移系列 $\mu = (\mu_1; \mu_2; \dots; \mu_T)$ を求めるために、次の値を定義する。ただし、 y_t は 1 拍の 1/480 を単位とする。また、 μ_t は各音に対応する状態で、1 から 12 の値をとる。

$$\pm_t(i) = \max_{\mu_1; \mu_2; \dots; \mu_{t-1}} P[\mu_1; \mu_2; \dots; \mu_{t-1}; \mu_t = i; y_1; y_2; \dots; y_t] \quad (4.3)$$

$\pm_t(i)$ は、 t 個目のイベントまでを観測した時点で、状態遷移が状態 i で終わっている状態遷移系列中のベストスコア（最も高い確率を与える状態遷移系列の確率）である。また、 \pm はモデルのパラメータを表す。ベストスコアの値は、次のような漸化式を満たす。ただし、次式において $A \bmod B$ は、 A を B で除算した余剰とする。 n は、End と Start を挿入した回数で、0 以上の整数である。

$$\pm_{t+1}(j) = \max_i [\pm_t(i) A_{ij}] \cdot b_j (y_{t+1} \bmod 480) \quad (4.4)$$

ただし、

$$A_{ij} = a_{ij} \quad (n = 0)$$

$$A_{ij} = a_{iE} \cdot a_{Sj} \cdot a_{SE}^n \quad (n > 0)$$

Viterbi アルゴリズムは、図 4.4 のようなトレリス上で、左から右に到る状態遷移系列の探索を行うアルゴリズムである。

このトレリスは、横軸にイベントを、縦軸に 1 から 12 の状態を配置しており、各イベントを 1 から 12 のいずれかの状態に対応づけてゆくことで、1 つの状態遷移系

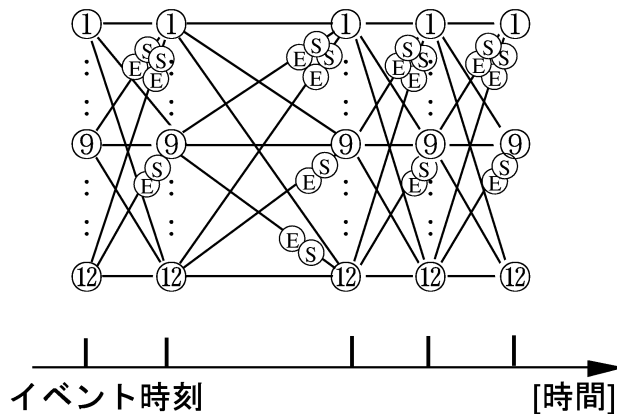


図 4.4: 最尤な状態遷移を求めるトレリス

列が得られる．拍線（拍の境目）をまたぐ状態遷移には End と Start を挿入する．このようなトレリス上で左から右に向かって式 (4.4) を用いながら $\mu_t(i)$ を計算してゆくことにより，最も事後確率の高い系列を選択することができる．

本研究ではテンポ一定の伴奏に合わせて弾いた演奏のクオンタイズを扱っているため，伴奏の拍線は既知だが，演奏の発音時刻はゆらいているため，演奏の拍線を見つけることは容易ではない．本研究では，演奏のイベント間隔とクオンタイズ後のイベント間隔との差を半拍以内にするという条件で，トレリス上で考えられるすべての2つのイベント間のパスで，拍線をまたいだか否かを判定する．実際の演奏でも，発音時刻間隔が半拍以上ずれることはほとんどないため，この判定条件は妥当であるといえる．ここでは，隠れ状態の拍中の時刻 q_t とクオンタイズ後のイベント間隔 $r_{t,t+1}$ という2つの概念を定義した上で，拍線をまたいだ回数を求める．

2 隠れ状態の拍中の時刻 q_t

q_t は，拍の頭を0としたときの，1から12までの各状態が対応するイベント時刻を示し次式で定義される．

$$q_t = \frac{\mu_t i}{12} \times 480 \quad (4.5)$$

2 クオンタイズ後のイベント間隔 $r_{t,t+1}$

$r_{t,t+1}$ は，2つのイベント間のパスが実際に選択されたと考えた場合のクオン

タイズ後のイベントの間隔であり，次式で定義される．

$$r_{t,t+1}(n) = (q_{t+1} \mid q_t) + n \times 480 \quad (4.6)$$

ここで n は，End と Start を挿入した回数である．

このとき，2つのイベントが拍線をまたいだ回数 \hat{n} は，実際の演奏のイベント間隔 $y_{t+1} \mid y_t$ とクオンタイズ後のイベント間隔 $r_{t,t+1}$ との差を最小にする n である (式 (4.7)) ．

$$\hat{n} = \underset{n}{\operatorname{argmin}} f(y_{t+1} \mid y_t) \mid r_{t,t+1}(n)g \quad (4.7)$$

たとえば，図 4.5 のような音符のクオンタイズを考えたとき，これに対応する適切な遷移は，状態 1 ($q_t = 0$) から状態 1 ($q_{t+1} = 0$) への遷移となる (実際には，あらゆる状態間の遷移を計算する) ．このとき，式 (4.6) で $q_{t+1} \mid q_t$ の値は 0 となり，代入すると式 (4.7) は

$$\hat{n} = \underset{n}{\operatorname{argmin}} f(y_{t+1} \mid y_t) \mid n \times 480g \quad (4.8)$$

となる．このとき， $1200 \cdot y_{t+1} \mid y_t \cdot 1680$ とすると \hat{n} の値は 3 となり，図 4.5 のように 3 回 End と Start を挿入すればよいと求まる．

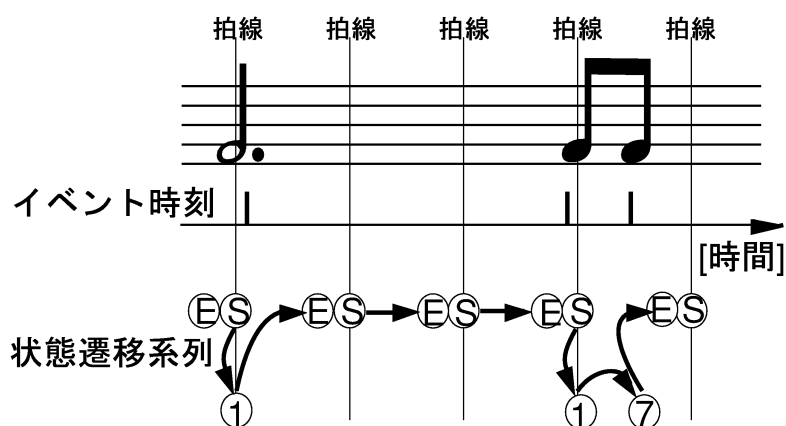


図 4.5: 付点 2 分音符の状態遷移の例

4つの HMM に分岐したモデルの場合には，各拍で 4つの HMM のどれを用いるかを，図 4.6 のような木構造を用いて探索していかなければならないが，プログラ

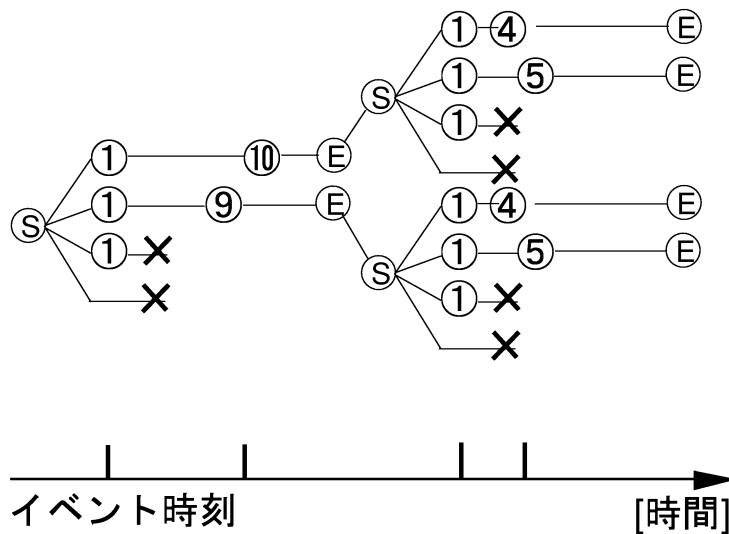


図 4.6: 4つのHMMによる探索木の例

ムとしては図 4.4 と同様のトレリス上の探索に帰着させることができる。その場合、トレリスの縦軸にあと 2 行を加え、16 分遷移型の状態 1 と、8 分 3 連遷移型の状態 1 と、頭のみ型の状態 1 とをそれぞれの別々の状態として扱う。また、音なし型の場合には、トレリス上に Start と End を 1 つずつ挿入する。

4.2.3 モデルパラメータの学習

発音時刻の遷移のモデル $P(\mu)$ すなわち a_{ij} と、発音時刻ゆらぎの順モデル $P(y|\mu)$ すなわち $b_j(k)$ を学習する。ここでは、教師つき学習と教師なし学習の両方を行った。

教師つき学習は、元々弾こうとした正規化された発音時刻系列 μ （正解データ）と、演奏された発音時刻系列 y の両方を学習用データとしてモデルパラメータを推定する手法である。教師つき学習によって得られたモデルパラメータを用いてクオンタイズすることにより、クオンタイズの基本的な性能を評価することができる。

一方、教師なし学習は、演奏された発音時刻系列 y のみを学習データとしてモデルパラメータを推定する手法である。実際に新たな演奏者の演奏に対してクオンタイズする場合には、教師データが与えられていないため、モデルパラメータを教師なし学習してクオンタイズする必要がある。

2 学習用データ

学習用のデータセットとして，人工的に生成したデータと，人間の演奏者のデータとの2通りを用意した．人工データは，プログラムの動作および，発音時刻のゆらぎの大きさによるクォンタイズの性能変化についての検証を意図したものである．

1. 人工データ

人工データの生成では，まずその拍を8分3連音符の演奏にするか，16分音符の演奏にするかを乱数で決定し，1拍に含まれる音数を1から 6^5 までの乱数で決定する．次に，それぞれの音が，どの位置に割り振られるかを乱数で決定する．8分3連音符では3通り，16分音符では，4通りのうちいずれかの位置となる．以上のようにしてできたランダムデータ μ_a の発音時刻を，平均0，標準偏差 $\frac{3}{4}$ の正規分布の乱数を用いて，ゆらぎをつけたものを人工データ y_a とする． $\frac{3}{4} = 10, 20, 30$ (1拍=480) とした3セットを作った．

2. 実演奏データ

3人の演奏者A, B, Cが，テンポ一定の伴奏に合わせてMIDIギターで即興演奏した記録を，実演奏データ y_h とする．演奏の長さは12コーラス(1コーラス=12小節)であり，そのうち約半分がソロ，残りが伴奏である．伴奏部分の多くは2音の和音の演奏である．各演奏者に2回ずつ，一方はテンポ120，もう一方は演奏者自身が設定したテンポでの演奏とし，合計6セットのデータを作成した．

3. 正解データ

学習を行うためには，正解データ，すなわち，演奏者が元々弾こうとした正規化された発音時刻を教師データとして与える必要がある．人工データの場合には，ゆらぎを加える前のランダムデータ μ_a が正解データとなる．

一方，人間の即興演奏データの場合には，正解データが存在しないので，人間が手作業でラベルづけをすることにより正解データ μ_h を作成した．具体的には，市販のシーケンスソフトウェア(Twelve Tone System, Cakewalk

⁵実験で使用した実演奏データの1拍内の音数の最大は6音であった．

Pro Audio 9) を用いて、ピアノロール表示⁶した音符の位置と長さを視覚的に確認しながら、1音ずつ適切な位置にラベルづけしていった。発音時刻のゆらぎが大きく、いくつかの正解候補が考えられた場合には、すべての候補のクオンタイズ結果を作って聴き比べ、より適切だと判断したものを採用した。

2 モデルパラメータの教師つき推定

正解データ μ と演奏された発音時刻系列 y から、 a_{ij} 、 $b_j(k)$ を推定した。出力確率 $b_j(k)$ の分布の学習は、学習データの音数の不足を補うため、音が入ったイベント時刻を中心とし標準偏差が5 (1拍=480) のガウス分布を足し合わせるにより求めている。

図 4.7 中の (a) は、人工データを作るときのゆらぎの標準偏差 $\frac{3}{4} = 20$ (1拍=480) とした場合の人工データから学習された、 $b_5(k)$ の分布である。(b) は、

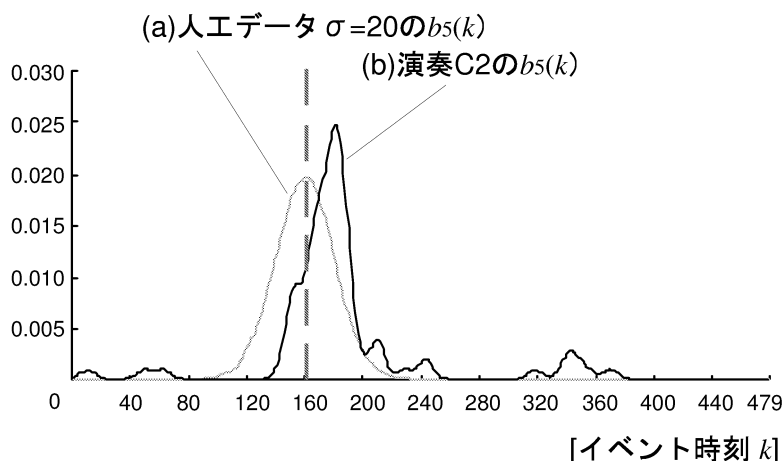


図 4.7: 人工、実演奏データから求めた $b_5(k)$

演奏者 C の 2 回目の演奏 (演奏 C2 と呼ぶ) から求めた $b_5(k)$ の分布である。人工データは、もともと正規分布の乱数でゆらぎを与えているので、 $b_5(k)$ の分布 (a) は、正規分布となった。一方、実演奏データから求めた、 $b_5(k)$ の分布 (b) もほぼ正規的な分布となったが、分布の中心がやや後ろにずれていた。このことは、状態 5 の位置の音符が、元々演奏者が弾こうとした発音時刻より

⁶ピアノロール表示とは、縦軸を音の高さ、横軸を時間とし、音の出るタイミングと鳴り続けている長さを表示するものである。

も遅れて演奏される傾向があることを示している．実際，演奏 C2 を調べたところ，8分3連音符の頭の音が通常より長く弾かれ，次の音の発音時刻が遅れる傾向が認められた．

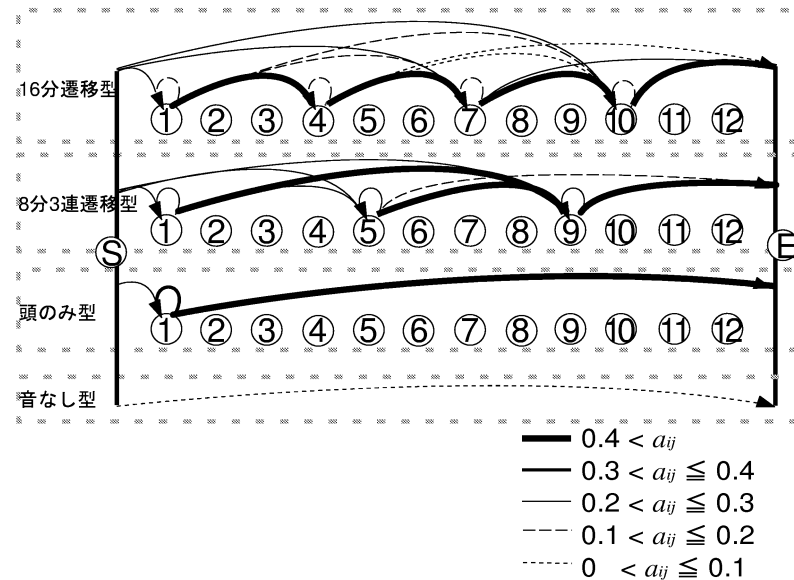


図 4.8: 演奏 C2 から求めた a_{ij}

図 4.8 は，演奏 C2 から求めた 4 つの HMM の遷移確率 a_{ij} である．8分3連遷移型，頭のみ型に対し 16分遷移型の自己ループの確率が低くなっていた．これは 16分遷移型では和音が出力されにくい傾向があることを示しているが，実際の C2 の演奏でもそのようになっていた．

2 モデルパラメータの教師なし学習

演奏された発音時刻系列 y のみから，その状態遷移系列の確率 $P(y) = P(y_{j_s})$ を最大化するようなモデルパラメータ θ_s を推定する．HMM では，隠れ状態遷移が観測できないため，パラメータを直接，最尤推定することはできない．そこで，Baum-Welch アルゴリズムを用いて θ_s および状態遷移系列の再推定を繰り返すことにより，尤度 $P(y_{j_s})$ を最大にするモデルパラメータ θ_s を求める．その際，学習データの不足を補うため，ヘルドアウト補間法 [北 99] を用いて $b_j(k)$ を補間する．

1. Baum-Welch アルゴリズム

Baum-Welch アルゴリズムを用いて、以下のように a_{ij} , $b_j(k)$ を推定する。推定に用いたのは、3人の演奏者 A, B, C が、テンポ一定の伴奏に合わせて MIDI ギターでジャムセッションした記録である。各演奏者 2 回ずつ、合計 6 セット (A1, A2, B1, B2, C1, C2) 用意した。

初期値の設定

a_{ij}^0 , $b_j^0(k)$ に適切な初期値を与える。 a_{ij}^0 は A1 から C2 までの 6 つの演奏から求めた a_{ij} の平均値とする。 $b_j^0(k)$ は、隠れ状態 i が対応するイベント時刻を中心とする分散 $\frac{3}{4} = 20$ の正規分布とする。 $b_j(k)$ が分散 $\frac{3}{4} = 20$ の正規分布に近い分布となることは、教師つき学習で確認している。

前向き、後ろ向き計算

モデル \mathcal{M} が与えられたとき、 $y_1; y_2; \dots; y_t$ を出力し、 t 番目のイベントで状態 i にいる確率 $\pi_t(i)$ (前向き確率) と、モデル \mathcal{M} と t 番目のイベントにおける状態 i が与えられたとき、 $t+1$ 番目以降に $y_{t+1}; y_{t+2}; \dots; y_T$ を出力する確率 $\bar{\pi}_t(i)$ (後ろ向き確率) とを算出する。

モデルパラメータ \mathcal{M} の更新

次のような漸化式で、 a_{ij} , $b_j(k)$ を更新する。

$$\begin{aligned}
 a_{ij}^{k+1} &= \frac{\text{状態 } i \text{ から状態 } j \text{ へ遷移する回数の期待値}}{\text{状態 } i \text{ から遷移する回数の期待値}} \\
 &= \frac{\prod_{t=1}^T \pi_t(i) a_{ij}^k b_j^k(y_{t+1} \bmod 480) \bar{\pi}_{t+1}(j)}{\prod_{t=1}^T \pi_t(i) \bar{\pi}_t(i)} \quad (4.9) \\
 b_j^{k+1}(k) &= \frac{\prod_{t=1}^T \pi_t(j) b_j^k(k)}{\prod_{t=1}^T \pi_t(j)}
 \end{aligned}$$

$$\begin{aligned}
& \prod_{s:t: (y_{t+1} \bmod 480) = k}^{t=1} \cdot \prod_{j=1}^{\infty} \mathbb{X}_{\text{t}(j)}^{-\text{t}(j)} \\
= & \frac{\prod_{s:t: (y_{t+1} \bmod 480) = k}^{t=1} \cdot \prod_{j=1}^{\infty} \mathbb{X}_{\text{t}(j)}^{-\text{t}(j)}}{\prod_{t=1}^{\infty} \cdot \prod_{j=1}^{\infty} \mathbb{X}_{\text{t}(j)}^{-\text{t}(j)}} \quad (4.10)
\end{aligned}$$

繰り返し

前向き，後ろ向き計算とモデルパラメータ θ の更新の計算をモデルパラメータが収束するまで繰り返す。

2. ヘルドアウト補間法

Baum-Welch アルゴリズムでは，一般に繰り返し計算をすることにより，尤度は単調に増加し，モデルパラメータは収束に向かうが，学習データのサンプル数が不十分な場合には， $\mathbb{X}_{\text{t}(i)}$ や $\text{t}(i)$ が 0 になる場合が多くなり， a_{ij} ， $b_j(k)$ の正しい推定ができなくなる．本研究では，音数の少ない状態 j でも $b_j(k)$ が正しく推定できるように，ヘルドアウト補間法を用いて， $b_j(k)$ を線形補間する．以下の説明では 1 から 12 までの各状態が対応するイベント時刻を時刻 0 として正規化した時刻 l を用いる (式 (4.11))．

$$\begin{aligned}
k_i - 40 \uparrow (j_i - 1) \geq 0 & \quad \text{のとき} \quad l = k_i - 40 \uparrow (j_i - 1) \quad (4.11) \\
k_i - 40 \uparrow (j_i - 1) < 0 & \quad \text{のとき} \quad l = k_i - 40 \uparrow (j_i - 1) + 480
\end{aligned}$$

$b_j(l)$ を以下のように線形補間する (式 (4.12))．補間係数 β は， $0 \leq \beta \leq 1$ である．

$$\hat{b}_i(l) = \beta \bar{b}(l) + (1 - \beta) b_i(l) \quad (4.12)$$

ここで， $\bar{b}(l)$ は，1 から 12 の状態に対応する各 $b_i(l)$ の平均の分布である (式 (4.13))．

$$\bar{b}(l) = \frac{\sum_{j=1}^{\infty} b_j(l)}{\sum_{j=1}^{\infty} 1} \quad (4.13)$$

式 (4.12) において $\hat{b}_i(l)$ は， β の値を適切に設定すると，最尤推定によって得られた $b_i(l)$ よりもよい推定値となることから，スタインのパラドックス

ス [EM77] として知られている．補間係数³を求める手法として，本研究ではヘルドアウト補間法 (推定法) を用いる．

ジェームズ・スタイン推定量 [EM77] では，補間係数 (縮減因子と呼ぶ) を求める方程式が定義されている．しかし，その方程式を使った場合， $b_j(l)$ ($j = 1; 2; \dots; 12$) の各分布で 0 から 479 までの各イベント時刻ごとに，補間係数を求めることになり，分布の形状が崩れていく (分布の山が無くなり平らな分布になっていく) という問題が起きる．そこで本研究では，補間係数³をすべての $b_j(l)$ ($j = 1; 2; \dots; 12; l = 0; 1; \dots; 479$) で同じとし，EM アルゴリズムにより最適な³を推定する．ジェームズ・スタイン推定量については，付録 E で述べる．

ヘルドアウト補間法は，言語モデルの 1 つである N グラムモデルのパラメータのスムージングにしばしば用いられる手法で，学習データ y を 2 つに分け，片方で最尤推定を行い，もう片方 (ヘルドアウト・データ) で補間係数³の推定を行う手法である．補間係数は EM アルゴリズムに基づいた繰り返し最適化により推定される．³の再推定式は，次式の通りである． T は³の再推定に用いた学習データの数である．

$$\hat{\lambda}^3 = \frac{1}{T} \sum_{t=1}^T \frac{\lambda^3 \prod b_j(l)}{(1 - \lambda^3) \prod b_j(l) + \lambda^3 \prod b_j(l)} \quad (4.14)$$

本研究では，学習データ y の前半分を Baum-Welch アルゴリズムによる最尤推定に使い，後ろ半分を補間係数³の推定に使う．ヘルドアウト補間法は，Baum-Welch アルゴリズムの再帰計算で $b_j(k)$ が推定されるごとに実行する．図 4.9 は演奏 A1 から教師つき学習と教師なし学習で推定した $b_4(k)$ の分布を比較したものである．両者は似通った分布となり，教師なし学習でも正しく出力確率 $b_j(k)$ が推定できていることが確認できる．

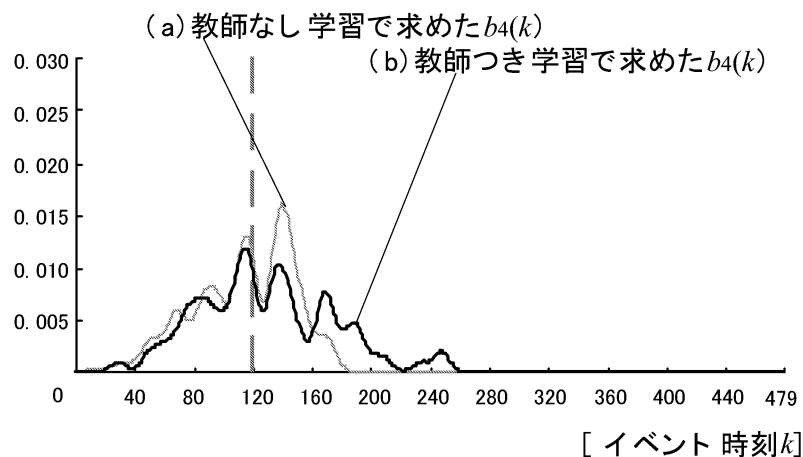


図 4.9: 教師つき/なし学習で求めた $b_4(k)$

4.2.4 クオンタイズ結果

ここでは、教師つき学習によって得られたモデルパラメータを使ってクオンタイズすることにより、モデルが正しい挙動をしていることを確認するとともに、教師なし学習で得られたモデルパラメータでクオンタイズすることにより、性能の確認を行う。

教師つき学習でのクオンタイズは、音単位と拍単位で評価し、教師なし学習でのクオンタイズは、音単位で評価した。音単位の評価は、各音の発音時刻が正解と一致している割合である。拍単位の評価は、拍の種類を、16分遷移型、8分3連遷移型、頭のみ型、音なし型の4種類に分けたとき、各拍の種類が正解の種類と一致している割合である。この4種類への分け方は、1つのHMMでのモデルの場合は、クオンタイズ結果から求め、4つのHMMに分岐したモデルの場合は、クオンタイズしたときに選択されたHMMの型から求めている。

表 4.1 に正解データに含まれる4種類の拍の割合を示す。演奏者 A, B が2回の演奏でそれぞれ似た傾向の演奏をしているのに対し、演奏者 C は、1回目は16分音符を多用した演奏、2回目は16分音符のまったくない演奏であった。

本研究では、一致率を以下のように定義し、クオンタイズの性能を評価する。

$$(\text{音単位の一一致率}) = \frac{(\text{正規化後の拍内位置が正解と一致した音数})}{(\text{正解データ中の音数})} \quad (4.15)$$

$$(\text{拍単位の一一致率}) = \frac{(\text{拍の種類が正解と一致した拍の数})}{(\text{正解データの拍の数})} \quad (4.16)$$

1. 機械クオンタイズ

まず、閾値処理による機械的なクオンタイズ（機械クオンタイズと呼ぶ）の性能を評価するために、8分3連、16分、16分3連の3種類の分解能で機械クオンタイズを実行し、音単位での一致率を求めた（表4.2）。尚、これ以外の分解能では一致率はさらに悪かった。

その結果、A1やC1、人工データのように16分の拍が多く含まれる演奏では、低い一致率しか得られなかった。ただし、C2のように16分で弾かれた拍がまったく無い演奏では、8分3連の分解能での機械クオンタイズが高い一致率を示すこともあった。

2. 制約つき機械クオンタイズ

実験に用いた演奏には、8分3連と16分の両方の音符が含まれている。したがって、1種類の分解能の機械クオンタイズで良い結果を得ることは困難である。そこでここでは、1拍内の音符が、8分3連系か16分系のいずれかであるという制約つきのクオンタイズを考える。制約つきクオンタイズでは、1拍内の発音時刻に対して8分3連および16分の2種類の機械クオンタイズを同時に行う。そして、クオンタイズ前の発音時刻の位置とクオンタイズ後の発音時刻の位置の距離（差の絶対値）の合計が少ない方のクオンタイズ結果を、拍ごとに選択していく。

その結果、人工データに対しては、 $\tau = 10$ では100%の音単位の一一致率を示し、 $\tau = 30$ でも96%以上という高い一致率を示していた。一方、実演奏データでは、A1とC1を除き、8分3連の機械クオンタイズより性能が劣っていた。このことから制約つきクオンタイズは、人工データのように演奏音の発音時刻の分布が正規分布でばらつきが小さい場合には有効であるが、多くの実演奏データなど演奏音のばらつきが大きい場合にはクオンタイズの選択が難しくなり、性能が低下することがわかった。

3. 学習に用いた演奏の教師つき HMM クオンタイズ

教師つき学習で獲得したモデルパラメータを用いて、学習に使ったのと同じデータについて HMM クオンタイズを実行し評価した(表 4.2)。以下に結果を述べる。

² 教師つき HMM クオンタイズと制約つき機械クオンタイズの比較

両者の一致率を比較したところ、実演奏データでは HMM クオンタイズのほうが高い値を示していた。このことから、発音時刻の状態遷移確率や出力確率に傾向(たとえば状態 5 の位置の音符が、元々演奏者が弾こうとした発音時刻よりも遅れて演奏される傾向)をもつ実演奏データの場合では、HMM によるモデルが有効であることが示された。本研究では、各状態での出力確率 $b_j(k)$ を 1 曲の学習データから得た分布を用いているため、曲の中で局所的にリズムがはしったり、もたったりした部分では、正しくクオンタイズできない場合があった。このような局所的なリズムの変化への対処については、今後検討していく必要がある。

一方、人工データでは HMM クオンタイズの性能が制約つき機械クオンタイズの性能を下回った。その原因としては、出力確率 $b_j(k)$ を学習する際に足し合わせたガウス分布の標準偏差の値が大きすぎたことが考えられる。標準偏差の値を大きくすると、学習データが少ない場合でも出力確率がなめらかになるという利点があるが、値が大きすぎると分布のすその広がってしまい、クオンタイズの性能が悪化するという欠点もある。人工データのように発音時刻の分布のばらつきが小さい場合には、標準偏差の値が小さいほうが高い性能が得られる。実際、人工データ ($\frac{3}{4} = 10$) を標準偏差を 1 として学習した出力確率 $b_j(k)$ で HMM クオンタイズしたところ、一致率は 100% となった。

² 1 つの HMM によるモデルと 4 つの HMM によるモデルの比較

両者の音単位の一一致率を比較すると、実演奏データでは、4 つの HMM によるモデルの方が、一致率が高かった。これは、状態 1 を、16 分遷移型、8 分 3 連遷移型、頭のみ型の 3 状態に分けることにより、各々の場合の $b_1(k)$ の分布を別々に学習し、より詳細にモデル化されたためであると考えられる。実際、拍単位の評価でも、4 つの HMM によるモデルのほう

が、1つのHMMによるモデルより高い一致率を示していることから、16分遷移型と8分3連遷移型の識別性能が向上したことが確認できた(表4.3)。

一方、人工データでは、1つのHMMによるモデルを使った場合と4つのHMMによるモデルを使った場合とが同じ一致率となった。これは、そもそも人工データが正規分布の乱数でゆらぎを与えているために、状態1を分けた場合の3つの $b_1(k)$ も、分ける前の $b_1(k)$ も、いずれも同じ正規分布となるためである。

2 クォンタイズ結果の例

図4.10は、演奏C2のクォンタイズ結果の一部を示したものである。機



図 4.10: クォンタイズ結果の例

械クォンタイズ(16分)では、8分3連音符のクォンタイズに失敗し、機械クォンタイズ(8分3連)では、16分音符のクォンタイズに失敗した。また、制約つきクォンタイズでも2拍目でクォンタイズに失敗した。これは、2拍目の8分3連の発音時刻が遅れ、16分の機械クォンタイズが選択されたためである。一方、HMMクォンタイズでは、1つのHMMと4つのHMMの両方とも、すべてのイベント時刻のクォンタイズに成功した。

表 4.1: 正解データに含まれる 4 種類の拍の割合

	人工データ			演奏者 A		演奏者 B		演奏者 C	
	$\frac{3}{4} = 10$	$\frac{3}{4} = 20$	$\frac{3}{4} = 30$	A1	A2	B1	B2	C1	C2
16 分の拍の割合	38.8%	38.8%	38.8%	21.4%	5.5%	6.5%	2.7%	37.2%	0.0%
8 分 3 連の拍の割合	37.8%	37.8%	37.8%	54.2%	68.7%	58.2%	85.8%	42.1%	39.7%
頭だけの拍の割合	6.9%	6.9%	6.9%	6.8%	10.1%	11.9%	6.4%	5.9%	26.7%
音がない拍の割合	16.4%	16.4%	16.4%	17.6%	15.7%	23.3%	5.1%	14.8%	33.8%

表 4.2: 機械クオンタイズと本手法の比較 (音単位の一致率)

	人工データ			演奏者 A		演奏者 B		演奏者 C	
	$\frac{3}{4} = 10$	$\frac{3}{4} = 20$	$\frac{3}{4} = 30$	A1	A2	B1	B2	C1	C2
機械クオンタイズ (8 分 3 連)	65.6%	58.2%	62.0%	67.6%	85.6%	79.4%	88.6%	57.0%	97.7%
機械クオンタイズ (16 分)	63.9%	67.9%	65.9%	54.5%	37.3%	36.8%	34.7%	70.7%	45.5%
機械クオンタイズ (16 分 3 連)	77.1%	70.7%	60.8%	57.7%	48.4%	57.8%	51.3%	56.1%	82.5%
制約つき機械クオンタイズ	100%	99.7%	96.3%	74.9%	66.0%	63.3%	48.1%	81.0%	77.6%
教師つき 1 つの HMM	99.6%	95.9%	86.5%	<u>75.9%</u>	<u>84.8%</u>	<u>80.0%</u>	<u>90.5%</u>	<u>85.1%</u>	95.0%
教師つき 4 つの HMM	99.5%	95.9%	86.5%	<u>82.3%</u>	<u>89.8%</u>	<u>81.4%</u>	<u>92.8%</u>	<u>85.5%</u>	95.7%
教師なし 1 つの HMM	99.6%	95.9%	86.2%	74.7%	78.4%	72.1%	78.2%	<u>84.7%</u>	89.4%
教師なし 4 つの HMM	99.5%	95.7%	85.4%	72.0%	76.2%	71.3%	76.5%	<u>82.2%</u>	94.7%

機械クオンタイズに対して、本手法のほうが性能が向上した箇所に下線をつけた。

表 4.3: 1 つの HMM によるモデルと 4 つの HMM によるモデルの比較 (拍単位の一一致率)

	人工データ			演奏者 A		演奏者 B		演奏者 C	
	$\frac{3}{4} = 10$	$\frac{3}{4} = 20$	$\frac{3}{4} = 30$	A1	A2	B1	B2	C1	C2
教師つき 1 つの HMM	98.6%	93.8%	81.3%	79.0%	85.4%	84.2%	91.9%	83.9%	94.9%
教師つき 4 つの HMM	98.6%	93.8%	81.3%	84.9%	89.2%	87.5%	93.5%	84.7%	97.3%

表 4.4: 他の演奏のモデルパラメータでのクオンタイズ結果

	人工データ			演奏者 A		演奏者 B		演奏者 C	
	$\frac{3}{4} = 10$	$\frac{3}{4} = 20$	$\frac{3}{4} = 30$	A1	A2	B1	B2	C1	C2
1) 自分ので別の演奏				75.9%	85.3%	79.4%	91.6%	55.5%	77.9%
2) 自分以外の演奏				70.2%	79.0%	59.2%	73.5%	76.9%	93.1%
3) ランダムデータで				73.3%	53.4%	60.8%	55.2%	82.5%	86.1%

4. 他の演奏のモデルパラメータでのHMMクオンタイズ

各実演奏データに対して、それ以外の演奏で学習したモデルパラメータを用いて、HMMクオンタイズを実行し評価した(表4.4)。評価に用いたモデルパラメータは、1) 同じ演奏者の別の演奏で学習したもの、2) 自分以外の2人の演奏で学習したもの、3) ランダムデータ($\frac{3}{4} = 20$)で学習したもの、の3通りである。評価は、4つのHMMによるモデルで行った。

1) は、ある演奏から得られたモデルパラメータが、同じ演奏者のほかの演奏に適用できるか調べるためのものである。1)の結果、演奏者A、Bは、その曲から学習したパラメータを使ってクオンタイズするのと比べて、一致率が最大6.4%しか劣っていなかった。一方、演奏者Cは、20%近く劣っていた。これは、2回の演奏をしたときに、演奏者A、Bはそれぞれ似た傾向の演奏を2度したのに対し、Cは、大きく異なる傾向の演奏をしたためだと考えられる。このことから、A、Bのように、演奏傾向が似ている場合には同じモデルで適用できるが、大きく異なる場合には、適用できないということがわかった。2)、3)は、モデルパラメータにいくつかの演奏の平均的なモデルや、正規的な分布を与えることにより、クオンタイズの性能がどう変化するかを調べたものであるが、いずれ場合も、A、Bでは1)より性能が劣っており、またCでも、1)より高くなったものの、その曲から学習したパラメータを使う場合と比べて一致率が低下していた。

このことは、本研究で提案したモデルが演奏者の発音時刻の特徴や癖を獲得しており、ある演奏者のモデルを一度作ればその演奏者が似た傾向の演奏をしたときには、同じモデルでクオンタイズが可能であることを示している。

5. 教師なしHMMクオンタイズ

教師なし学習したモデルパラメータを用いて、クオンタイズを実行し評価した(表4.2)。

教師なしの結果は、人工データでは、教師つき学習の結果とほぼ同じになった。これは、人工データでは、学習データが充分多く、また、各状態の出力分布の重なりが少ないため、出力分布が正しく求まったためである。

一方、実演奏データでは、教師なしの結果は、教師つきの結果よりは若干低い

一致率となったが、すべての演奏で7割以上となり、教師なしが高い性能を示すことが確認できた。4つのHMMによるモデルは1つのHMMによるモデルに比べて低い一致率となった。これは、4つのHMMでは、状態数が増加したために、モデルの教師なし学習がより困難になったためである。したがって、ジャムセッションの演奏記録など、限られた長さの演奏をクオンタイズする場合には、1つのHMMによるモデルで教師なし学習したモデルパラメータでのクオンタイズが有効である。

4.3 フレーズ分割

本節では、楽曲をある短いまとまりを持った長さ（グループ）に自動で分割する。多声音楽のグルーピングでは、音高方向と時刻方向の2通りのグルーピングがある。音高方向のグルーピングは、同時に進行する複数の旋律の分割で、たとえば、ピアノの左右それぞれの手に対応する演奏の分割や、複数のパートの分割である（図4.11.(a)）。また、時刻方向の分割は、演奏の時間的な切れ目の分割である（図4.11.(b)）。このような、グルーピングの自動化が実現すれば、分割されたグループをフレーズデータベースに蓄積し、演奏者固有のフレーズの模倣が可能となるばかりでなく、音楽構造の階層的な解析 [平田 02] への利用も期待できる。

(a) 音高方向のグルーピング

Allegretto



(b) 時間方向のグルーピング



図 4.11: グルーピングの例

楽曲のグルーピングに関する理論としては、グループを分割するための複数のルールを定義した Generative Theory of Tonal Music (GTTM) [LJ83] がある。しかし、GTTMのグルーピング構造分析では単旋律の楽曲が対象で、和声に関する分析が欠如しているため、我々が目標とする多声音楽のグルーピングには対応していない。そこでここでは、GTTMの規則を直接計算機上の実装するのではなく、別の手法により自動グルーピングを実現し、その結果が GTTM の結果と近くなるようにする。ここでは、そのような手法としてボロノイ線図を用いたグルーピングを提案する。以下、4.3.1 では GTTM のグルーピング構造分析について概説しその実装の困難さについて述べ、4.3.2 では、これまで実装された、GTTM 以外のグルーピング手法について述べる。4.3.3 ではボロノイ線図によるグルーピング手法を提案し、その性能を評価する。4.3.4 では、実際にボロノイ線図を用いてフレーズを分割し、手作業による分割と提案手法とを比較する。

4.3.1 Generative Theory of Tonal Music

GTTM は、人間の音楽知覚・認知過程をもとに、楽曲の階層的な構造の分析を目的としたもので、分析結果は、音の重要度、楽曲の階層構造を木構造で表現する。GTTM はグルーピング構造分析、拍節構造分析、タイムスパン簡約、プロロンゲーション簡約という4つのサブ理論からなるが、ここでは、グルーピング構造分析について説明する。グルーピング構造分析は、楽曲をより小さい構造(グループ)に分割するとともに、各グループの階層構造を決定する。解析ルールには、グルーピング構成ルール GWFR (Grouping Well-Formedness Rules) とグルーピング選好ルール GPR (Grouping Preference Rules) とがある。

グルーピング構成ルール GWFR

グルーピング構成ルールは、グループの構造を定義したルールで、以下の5つが定義されている。

GPFR1: 構成要素が連続している場合のみグルーピングを形成できる。

GPFR2: 1つの曲は1つのグループである。

GPFR3: グループはより小さなグループ(サブグループ)を内部に含んでもよい.

GPFR4: グループがサブグループの一部だけを含むことは許されない. サブグループ全体を内部に含まなければならない.

GPFR5: グループがサブグループを含むなら, グループ構造が交差しないサブグループで内部を埋めつくさなければならない.

図 4.12.(a) は, 正しいグルーピングの例, 図 4.12.(b) は誤ったグルーピングの例である.

(a) 正しいグルーピングの例



(b) 誤ったグルーピングの例 グループ構造が交差 (GWFR5 が不成立)



サブグループの一部のみ含む (GWFR4 が不成立)
構成要素が連続していない (GWFR1 が不成立)

図 4.12: 正しい/誤ったグルーピングの例

グルーピング選好ルール GPR

グルーピング選好ルールは, 好ましいグルーピングを作るルールで, 以下の7つが定義されている.

GPR 1 から GPR3 は, 局所的なレベルにおけるルールである.

GPR1 alternative form: 非常に小さいグループへの解析は避ける. 特に単音をグループとすることは避ける.

GPR2 proximity: 連続した4音 (n_1, n_2, n_3, n_4) において, 以下の条件のいずれかが成立する場合, n_2 と n_3 の間がグループの境界と認識される (n に休符は含めない).

a.slur/rest n_2 の終わりから n_3 の始まりまでの時間間隔が, n_1 の終わりから n_2 の始まりまでの時間間隔および, n_3 の終わりから n_4 の始まりまでの時間間隔よりも長い.

b.attack-point n_2 の始まりから n_3 の始まりまでの時間間隔が, n_1 の始まりから n_2 の始まりまでの時間間隔および, n_3 の始まりから n_4 の始まりまでの時間間隔よりも長い.

GPR3 change: 連続した4音 (n_1, n_2, n_3, n_4) において, 以下の条件のいずれかが成立する場合, n_2 と n_3 の間がグループの境界と認識される (n に休符は含まない).

a.register n_2 ; n_3 間の音高差が n_1 ; n_2 間の音高差および n_3 ; n_4 間の音高差よりも大きい.

b.dynamics n_2 ; n_3 間でダイナミクスの変化があり n_1 ; n_2 間, n_3 ; n_4 間ではそれがない.

c.articulation n_2 ; n_3 間で, アーティキュレーションパターンの変化があり n_1 ; n_2 間, n_3 ; n_4 間ではそれがない.

d.length n_2 と n_3 が異なった音長をもち, n_1 と n_2 もしくは n_3 と n_4 が同じ音長である.

GPR4 から GPR6 は, 大局的なレベルにおけるルールである.

GPR4 intensification: GPR2, GPR3 がより多く成立する境界が, 大局的なレベルでもグルーピング境界になる可能性が高い.

GPR5 symmetry: グループを2等分するグルーピングを優先する.

GPR6 parallelism: 2つ以上の並列した部分 (parallel) を含むようなグルーピングを優先する.

GPR7 は, グルーピング構造分析とタイムスパン構造分析, プロロンゲーション簡約との関係におけるルールである.

GPR7 Time-Span and Prolongational Stability: タイムスパン簡約, プロロンゲーション簡約がより安定するグルーピング構造を優先する.

GTTM グルーピングルール実装上の問題

GTTM は、音楽に関して専門知識のある聴取者の直観を形式的に記述するための理論であり、その構成の緻密さから計算機上への実装に適していると考えられている。しかし、GTTM のグルーピング構造分析によるグルーピングの自動化は、以下のような問題があり、実現は困難である。

問題点 1：多声音楽が直接扱えない

グルーピング構造分析は、単旋律の楽曲が対象なため、多声音楽を直接分析することはできない。文献 [井田 01] では、2 つの単旋律からなる楽曲を対象に、GPR2, 3 の計算機上への実装を試みたが、ヒューリスティックスを多用しており、制限された条件下で動作するものであった。また文献 [東洋 02] では、多声音楽を対象に、あらかじめヒューリスティックスにより隣接する 4 音が声部進行に成り得るか判断した上で、GPR2, 3 の適用を行った。これら [井田 01, 東洋 02] の手法におけるグルーピング結果（特に、音高方向のグルーピング結果）は、ヒューリスティックに依存していると考えられる。

問題点 2：定義の曖昧性

GTTM のルールは定量的に定義されておらず、複数のルール間の優先順位が不明確で競合が起きる。文献 [Del87] では、GPR の知覚的強さに関する実験的研究が行われ、文献 [PSR93, SP94] では、その成果をもとに、GPR の各ルールに重みをつけ、単旋律の楽曲を対象として、フレーズ分割を行った。しかし、そこで実装できたのは、GPR2, 3 のみで、GPR4 以降の大局的なグルーピングは実装できていない。これは、GPR4 以降では、「可能性が高い」、「優先する」という曖昧な用語により規則が定義されているために、計算機上での取り扱いが困難なためである。

問題点 3：解のフィードバック

GTTM では、グルーピング構造分析の分析結果を、拍節構造分析、タイムスパン簡約、プロロンゲーション簡約に用いるだけでなく、逆にタイムスパン簡約、プロロンゲーション簡約の分析結果をフィードバックしてグルーピング構造分析に用いている。したがって、グルーピング構造分析を完全に実装するためには、GTTM のその他のサブ理論も実装しなければならない。これまで、

タイムスパン簡約の実装 [Hal94] やプロロンゲーション簡約の実装 [上符 98] を試みた例もあるが、すべてのサブ理論を統合したシステムは実現していない。

4.3.2 GTTM 以外のグルーピング手法

GTTM 以外のグルーピング手法で、これまで実装されたものとしては、ルールベースに基づくシステム [AW97, Tem01, ST]、類似度に基づくシステム [Wey02]、不連続性に基づくシステム [Row92, Cam01] などがある。

ルールベースに基づくシステム

文献 [AW97] では、メロディをグループに分割して、それを分類するシステムを実装した。しかし、そこでのグルーピングはメロディを時間方向で等間隔に分割するものであった。文献 [Tem01, ST] で実装された Melisma は、ルールベースにより、拍節認識、グルーピング、コード認識、調性認識などを自動で行うシステムであった。グルーピングの仕組み自体は、2音間の IOI(inter-onset interval) と OOI(onset-to-onset interval) を計算し、その距離の大きさに応じて、ペナルティを与え、ペナルティが閾値を越えると、グループの境界とするものであった。これらのシステム [AW97, Tem01, ST] は、いずれも単旋律を対象としているため、和音を含むジャムセッションの演奏記録の分割に使うことはできなかった。

類似度に基づくシステム

文献 [Wey02] で実装された ISSM(Integrated Segmentation and Similarity Model) は、メロディの中から類似な部分を見つけながら、グルーピングを行うシステムであった。したがって、ポピュラーミュージックのように、同じメロディが繰り返されるような演奏では、有効な手法であったが、ジャムセッションのように新たなメロディが次々と入力される演奏のグルーピングには向いていなかった。

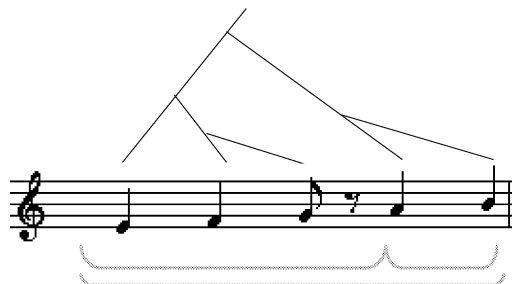
不連続性に基づくシステム

文献 [Row92] の Cypher では、音域や音量、音長、和声、拍位置などをリアルタイムで検出し、その不連続性を数値化し、それが閾値を超えた場合にフレーズの境界とした。文献 [Cam01] で実装された LBDM(Local Boundary Detection Model) では、2 音間の IOI, OII, 音高差を求め、その変化の大きい部分をグループの境界とした。これらのシステム [Row92, Cam01] で実現できたのは、GPR2(proximity), GPR3(change) に相当する局所的なグルーピングで、GPR4 以降に相当するような、大局的なグルーピングは実現していない。

4.3.3 ポロノイ線図に基づくグルーピング

GTTM を直接計算機上に実装することは困難であり、ここではポロノイ線図を用いたグルーピング手法を提案する。GTTM が、音の重要度、階層構造を木構造で表

(a) GTTMによる表現



(b) ポロノイ線図 (ベン図) による表現

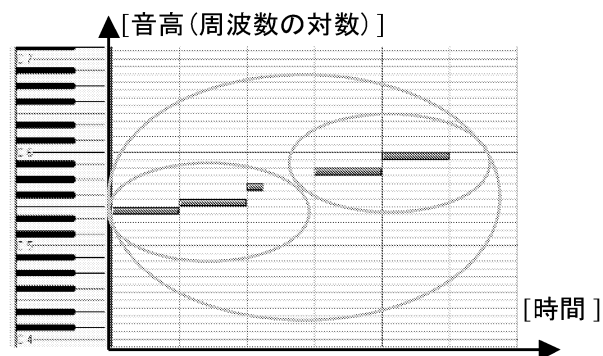


図 4.13: GTTM とベン図によるグルーピングの表現

現し、音符を楽譜で表現していたのに対し(図4.13.(a)), ボロノイ線図を用いたグルーピングでは、階層構造をベン図で表現し、音符はピアノロール⁷で表現する(図4.13.(b))。

ボロノイ線図

ボロノイ線図は、ある空間内に幾つかの図形成分が与えられた時に、最近傍則によって、その空間を排反な部分空間に分割したときに生じる分割境界の線図形である[浅野90]。2次元空間内に n 個の点を与えられた場合のボロノイ線図は、2点間の垂直2等分線をつなげ合わせた形状となる(図4.14)。

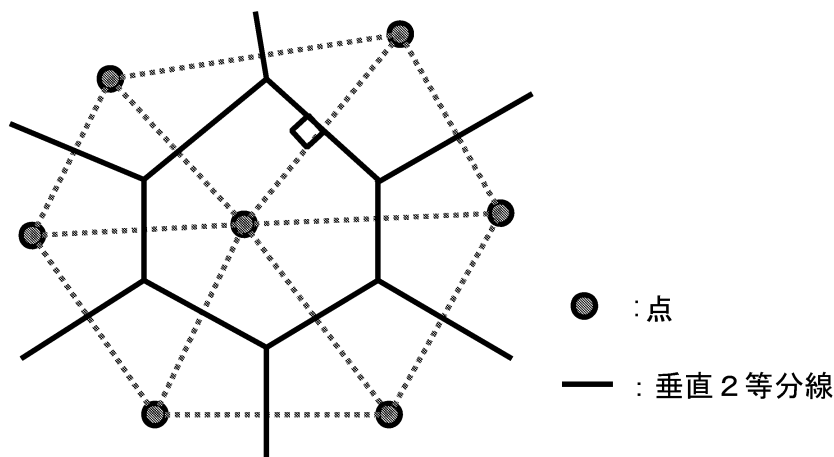


図4.14: n 個の点によるボロノイ線図

水平な2直線によるボロノイ線図

ピアノロール上の音符は時間軸に水平な直線で表される。したがって、そのボロノイ線図を考える場合には、水平な複数の直線により生じるボロノイ線図を書く必要がある。図4.14のような n 個の点によるボロノイ線図は複数の直線により構成されていたが、2本の直線によるボロノイ線図は、直線と2次曲線により構成される。ここではまず、2本の水平な直線により生じるボロノイ線図について説明する。簡単のため、縦軸を y 軸、横軸を x 軸、現在の時刻を t とする。そして、2本の直線を

⁷ピアノロールとは縦軸を音の高さ、横軸を時間とし、音の出るタイミングと鳴り続けている長さを表示するものである。

a と b とする。 a, b が時間的に重なっている部分，すなわち，2つの音が同時に鳴っている部分でのボロノイ線は， $x = t$ と a との交点と $x = t$ と b との交点の2等分点の集合となり，2本の直線の垂線を2等分するような直線となる（図 4.15.(a)）。 a だけの部分では， $x = t$ と a との交点から一番近い b 上の点と a との交点の集合となり，2次曲線となる（図 4.15.(b)）。 図には無いが， b だけの部分がある場合には，それも同様に2次曲線となる。 a, b どちらも通っていない場合には， $x = t$ から一番近い a 上の点と， $x = t$ から一番近い b 上の点との等距離にある点の集合で直線となる（図 4.15.(c)）。

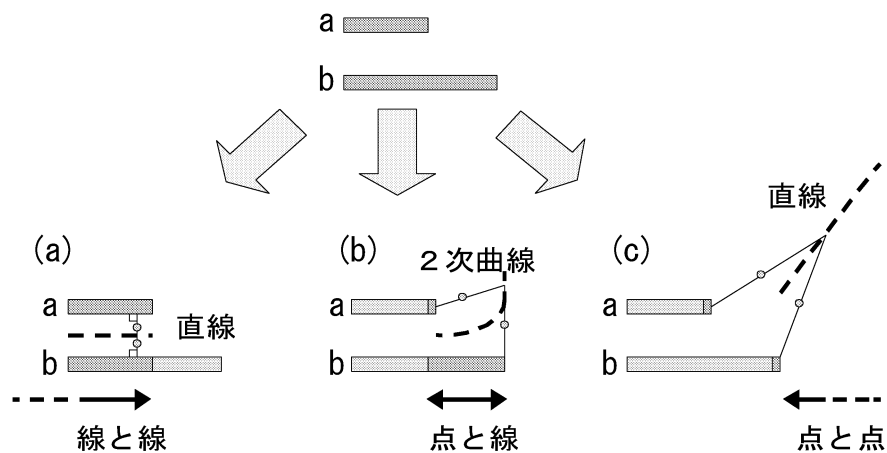


図 4.15: 水平な2直線によるボロノイ線図

3本以上の直線によるボロノイ線図

3本以上の水平な直線のボロノイ線図を描く場合には，すべての直線間のボロノイ線をすべて書き，不要な部分を削除する。たとえば，直線 a, b, c によるボロノイ線図を書く場合には，まず3本のボロノイ線をすべて描く。すると，それらのボロノイ線は a, b, c から等距離の点で交わる。そして，直線 a と直線 b によるボロノイ線は，交点を境に，片方は a, b よりも c の近くを通るため，その部分は削除する。すなわち， a, b のボロノイ線図のうち， c の領域に入ったものは削除する（図 4.16）。 以上のような処理を行うことにより，2本以上の水平な直線によるボロノイ線図を書くことができる。

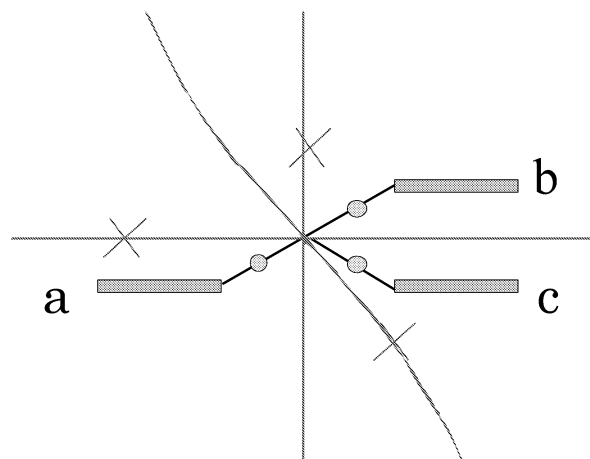


図 4.16: 2 本以上の直線によるポロノイ線図

ピアノロール上のポロノイ線図の問題

ポロノイ線図をピアノロール上に描く場合には、縦横のスケールリングが問題となる。すなわち、ピアノロールの縦軸は音高(半音)、横軸は時刻(tick)を表しており、スケールリングする必要がある。スケールリングの比率により、ポロノイ線図の形状は変化する。本研究では、1半音 = 90tick、1半音 = 100tick、1半音 = 110tick の3通りのスケールリング比で実験を行った。

グループの結合

以上のようなポロノイ線図による分割を、関係の近い音から順に結合していくことにより、階層的なグルーピングができる。ここでは、グルーピングするためのルールとして、以下の2つを導入する。

- 2 ポロノイ線図により囲まれた面積の小さなグループから順に結合する。
- 2 結合するグループは、そのグループに最も近い位置にあるグループに結合する。

ポロノイ線図に囲まれている部分の面積の計算は、内部に含まれる格子点の数で代用する。格子点は音高 36 から 96 までの範囲で打ち、その間隔は 10tick ごとである。たとえば、1半音 = 110tick とした場合には、縦に並ぶ格子点の数は $(96 - 35) \times 110 \div 10 = 671$ 個である。

グルーピング結果

ポロノイ線図を用いたグルーピングを計算機上に実装した．提案手法のグルーピング結果 (図 4.17) と GTTM に基づき手作業で作成したグルーピング結果 (図 4.18)⁸とを比較した．

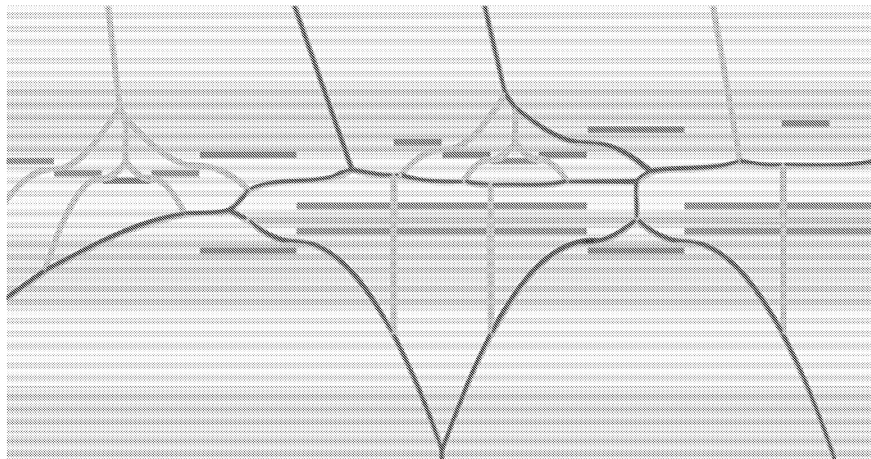


図 4.17: ポロノイ線図に基づくグルーピング結果

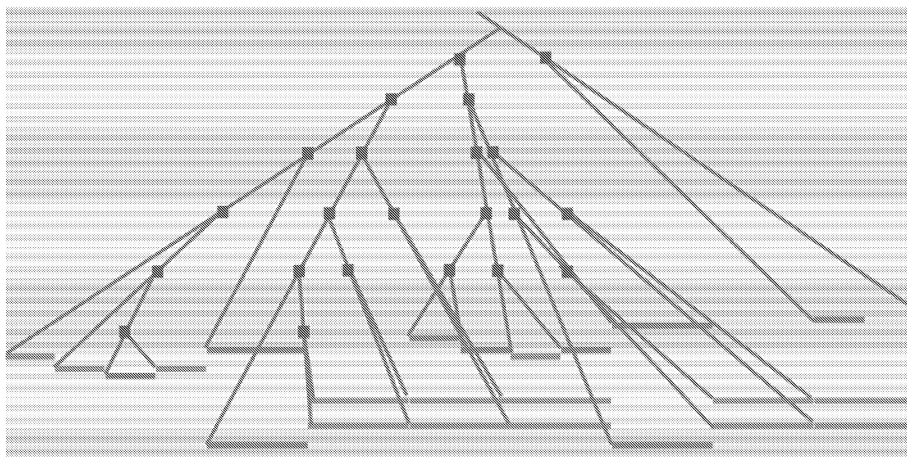


図 4.18: GTTM に基づき手作業によるグルーピング結果

⁸図 4.18 は，文献 [平田 02, HM02] に基づき GTTM のタイムスパン簡約を多声音楽に拡張したものである．

2 定性的な結果

1) 発音時刻と音高差に関するルール

GTTMでは、発音時刻と音高差に関するルールがGPR2とGPR3.aの2つに分かれており、ルールを適用する際にしばしば競合が起こったが、提案手法では、ポロノイ線から音までの距離という1つの基準に統一したため、競合が起これなくなった。ただし、提案手法では、音高と時刻のスケーリング比率を適切に指定する必要がある。

2) グループの長さに関するルール

GPR5(等しい長さのグループを優先させる)は、提案手法では、小さい大きさのグループから結合していくことにより結果として均等な長さのグループが実現できた。

3) その他のGTTMルール

GPR1(小さなグループへの解析は避ける)は、提案手法では、面積の小さなグループから結合することにより実現できた。GPR3.b,d(ダイナミクス,音長に対するルール)は、今回は実現できなかった。

2 定量的な結果

式(4.17)に示す一致率を用いて、3種類のスケーリング比でグルーピングの性能を定量的に評価する(表4.5)。ポロノイ線図で囲まれた図形が結合する時、それらがGTTMのタイムスパン木上で親子関係にあるある場合を正しいグルーピング、そうでない場合を正しくないグルーピングとした。実験に用いたのは、「トルコ行進曲」のMIDIデータである。

$$(\text{グルーピングの一致率}) = \frac{(\text{正しくグルーピングできた音数})}{(\text{すべての音数})} \quad (4.17)$$

実験の結果すべてのスケーリング比で70%以上の一致率となった。

表 4.5: 3つのスケーリング比におけるグルーピング結果

スケーリング比 (ticks/半音)	90	100	110
一致率	78.5 %	78.5 %	72.8 %

4.3.4 ボロノイ線図に基づくフレーズ分割

以上のようにしてできた，グルーピング結果のうち，1つの階層を選択することによりフレーズ分割を行う．ボロノイ線図を用いたグルーピングも，GTTMによるグルーピングと同様に階層構造を持っている．すなわち，グルーピングの一番上の階層では，すべての音がひとつのグループに属し，階層が下になるにつれてグループの数が増加する．したがって，フレーズ分割の際に，どの階層を選択するかによって，各フレーズの長さが変化する．比較的下にある階層を選択した場合には，1つ1つのフレーズの長さが短くなる．そしてその結果，フレーズを再利用して新たな演奏を生成するとき，フレーズの組み合わせの数が大きくなり，バリエーションの豊富な演奏となる．細かく分割されたフレーズから質の高い演奏が生成可能であることは文献 [Cop92] に示されている．しかし，フレーズを細かく分割しすぎると，フレーズに含まれている演奏者固有の特徴が無くなり，個性のないフレーズになる恐れがある．フレーズ分割をする際には，適切な階層を選択する必要がある．

フレーズ分割

ジャムセッションの場合，多くのフレーズは長さ1小節のフレーズを組み合わせで演奏しているため，多くのフレーズが1小節前後の長さになるような階層が適切であると考えられる．そこで本研究では，各グループの長さが1小節程度になるよう，すべてのグループの平均の長さが1小節にもっとも近くなる階層を選択し，フレーズ分割を行うことにした．

分割結果

フレーズ分割の性能を，以下の一致率で評価した．

$$(\text{フレーズ分割の一致率}) = \frac{(\text{フレーズが正解データと一致した数})}{\text{正解データのフレーズの数}} \quad (4.18)$$

評価に用いたのは，4.2.4でクォンタイズ済みの3人の演奏から抜き出した100小節である．正解データは手作業で作成した．評価するのは，(1)ボロノイ線図に基づきフレーズ分割した結果と(2)演奏記録を小節線で機械的に分割した結果，である(表4.6)．スケーリング比は100とした．

演奏者 A,B では (1) が (2) より高い性能を示した。一方、演奏者 C では (2) のほうが高い性能を示した。演奏者 C で (1) が不正解となった部分の演奏の多くは、伴奏のように同じフレーズを繰り返し演奏する箇所であった。このような部分では、小節線による分割のほうが有効となる場合もあったが、多くの場合ではポロノイ線図によるフレーズ分割が有効であることが確認できた。

表 4.6: 本手法と小節線での分割結果の比較

	演奏者 A	演奏者 B	演奏者 C
(1) ポロノイ線図に基づく分割	71.5 %	70.6 %	74.1 %
(2) 小節線による機械的な分割	68.4 %	55.2 %	78.5 %

4.4 フレーズの再利用

ここでは、切り出したフレーズを意図空間上に配置し、再利用する手法として、MDA-OR によるフレーズの空間配置法、局所自己相関関数によるフレーズからの特徴抽出法、正準相関分析による重要な物理特徴量の選択法について述べる。

4.4.1 意図空間の再構成

3.3.3 で述べたように、意図空間は、一対比較によって求めたフレーズ間の主観的類似度に基づきフレーズを配置した空間である。そして、意図空間と出力演奏の物理的な特徴との相関は正準相関分析により求めていた (図 4.19)。このときフレーズのサンプル数が多いほうが、より密な写像を求めることができる。一方、3.3.2 で振る舞いのモデルを学習する際の意図空間の次元は、振る舞いのモデルの学習する上でできる限り低いほうが好ましい。

しかし、この2つを同時に満たすことは困難である。なぜなら、フレーズを空間に配置する際にはフレーズ数が増加するに従って、低次元で配置することが一般に困難だからである。たとえば、類似度データを空間に配置する一般的な手法として、Kruskal の多次元尺度法 [KW78] がある。Kruskal の多次元尺度法は、2つのフレー

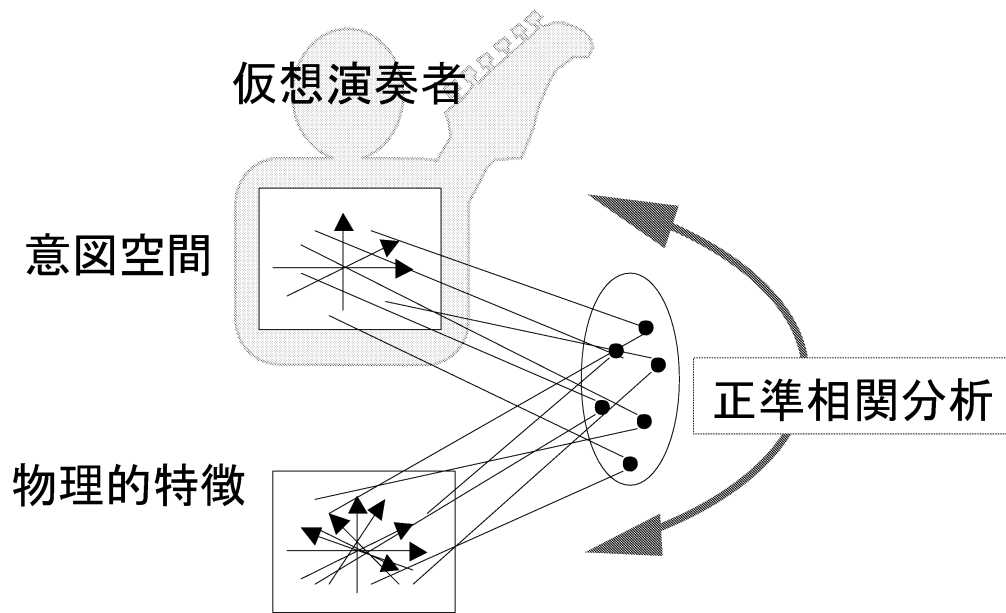


図 4.19: 意図空間と出力演奏の物理的な特徴との相関

ズ j と k の類似度を \pm_{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 (4.19)) .

$$\pm_{jk} > \pm_{lm} \text{ならば } d_{jk} \cdot d_{lm} \quad (4.19)$$

このとき, 式 (4.19) が成立する度合いはストレス値 S で表され, S の値が小さいほど類似度をよく反映した空間となる. S の値はデータ数が増加するにしたがって増加し, また, 次元数が少なくなるほど増加するため, このような手法で, 我々が求めるような空間を構成することは困難である. これは Kruskal の手法が類似度の順序関係を保存しようとするが, データの数が増えるにしたがってそこに無理が生じているためである.

この問題を解決するため, 本研究では, フレーズの空間への配置を多次元尺度法の1つである MDA-OR (Minimum Dimension Analysis of Ordered Class Belonging) [林 88] を用いて行う (付録 F 参照). MDA-OR は, 式 (4.19) のような順序関係を完全に成立させることを考えるのではなく, その関係の成立する比率を大きくすることを考えたものである. 被験者実験で求めた類似度のように, 数量そのものが厳密な意味

をもつものではない場合には、漠としたものを漠と扱い、要素の大局的な空間配置をする MDA-OR の手法が適切だと考えられる。

4.3.3 で分割した 100 個のフレーズ間の主観的類似度を一対比較により求め、階層的次元作成法により適合度をみながら MDA-OR の次元を上げていった結果、3 次元で適合度が 99.2 となり充分精度が高くなったと考え、空間の次元を 3 次元に決定した。3 軸のうち、最も重要な次元は 1 軸で、次に重要なのが 2 軸である。図 4.20 は、100 個のフレーズの Kruskal の多次元尺度法による空間配置と、MDA-OR による空間配置とを比較したものである。Kruskal の手法では、フレーズがほぼ一様な分散で配置されているのに対し⁹、MDA-OR では、1 軸の方向で大きく 2 つのグループに分かれていた。

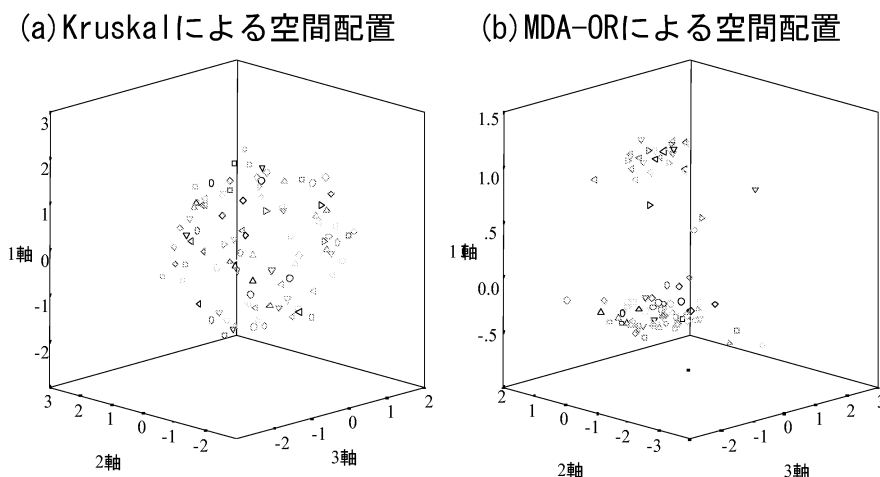


図 4.20: Kruskal と MDA-OR による空間配置

2つのグループのフレーズを調べた結果、それぞれソロのようなフレーズのグループと伴奏のようなフレーズのグループであることがわかった。このことは、被験者が類似度判定する上で、フレーズがソロであるか伴奏であるかが重要であるということを表している。本ジャムセッションシステムでは、ソロのフレーズと伴奏のフレーズとの区別はしておらず、仮想演奏者がソロのような演奏を弾くか伴奏のような演奏を弾くかはそのときの意図ベクトルによって決定される。

⁹Kruskal の手法では、図 4.20(a) のように、球の表面にフレーズが並んだようになることがしばしばある。これは、無理に低次元化したためで、フレーズが適切に配置できていないことを示している。

4.4.2 局所自己相関関数を用いた特徴の抽出

局所自己相関関数を用いて、フレーズの物理的な特徴を抽出する。ジャムセッションでは、あるフレーズを音高方向、時間方向に平行移動したようなフレーズが使われることがある。特に、伴奏では、コード進行に応じてフレーズ全体の音高を変化させることが多い。このような、あるフレーズとそのフレーズを音高方向や時間方向

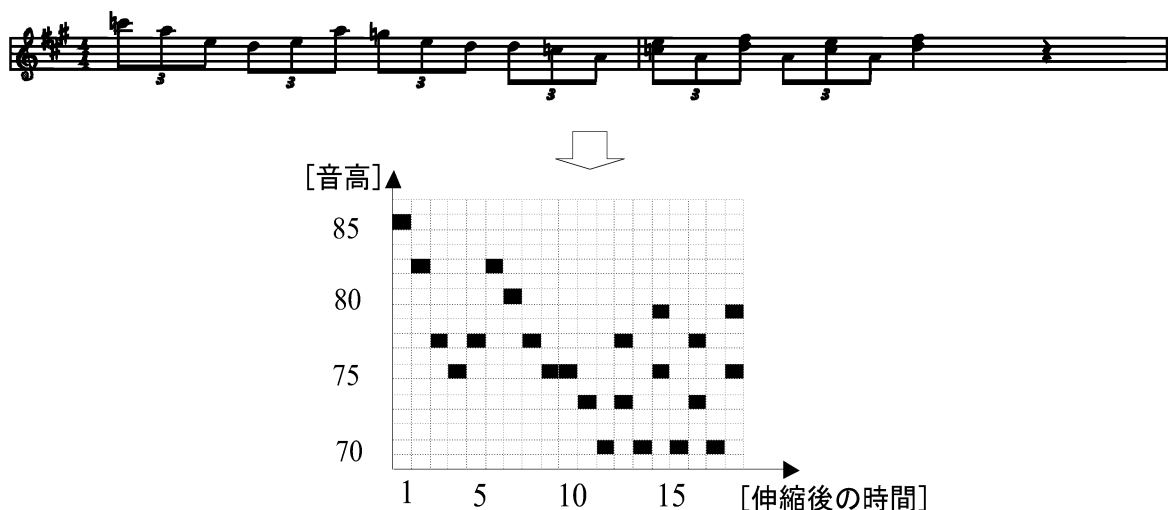


図 4.21: 時間軸を伸縮させたピアノロール

に少しずらしたフレーズは意図ベクトルが類似していると考えられる。したがって、その2つのフレーズからは、同じような特徴量が抽出されることが好ましい。そのようなことを実現する方法として、発音時刻が等間隔となるように時間軸を伸縮したピアノロール上のフレーズ(図 4.21)に対して局所自己相関関数に基づく特徴を抽出した。自己相関関数は平行移動に対して不変であることが知られているが、その高次への拡張が、N 次局所自己相関関数である [MR68, 大津 81]。ピアノロールの横軸を伸縮したのは、音の長さよりも、音高の変化が特徴として重要だと考えたためである。ピアノロールの縦軸は、ギターの音域を考え、MIDI のノートナンバー 36 から 96 までとする。

このようなピアノロール上で音が出ているマスをも $f(r) = 1$ 、音が出ていないないマスをも $f(r) = 0$ とすると N 次自己相関関数は、変位方向 $(a_1; \dots; a_N)$ に対して、

$$x^N(a_1; \dots; a_N) = \int f(r)f(r+a_1)\dots f(r+a_{N-1})f(r+a_N)dr \quad (4.20)$$

で定義される．従って， N 次局所自己相関関数は，次数や変位方向のとり方により無数に考えられるが，ここでは次数を1，局所領域を 5×5 に限定する．局所領域の平行移動により同じになる特徴を除くと，特徴の数は全部で13個になる（図4.22）．

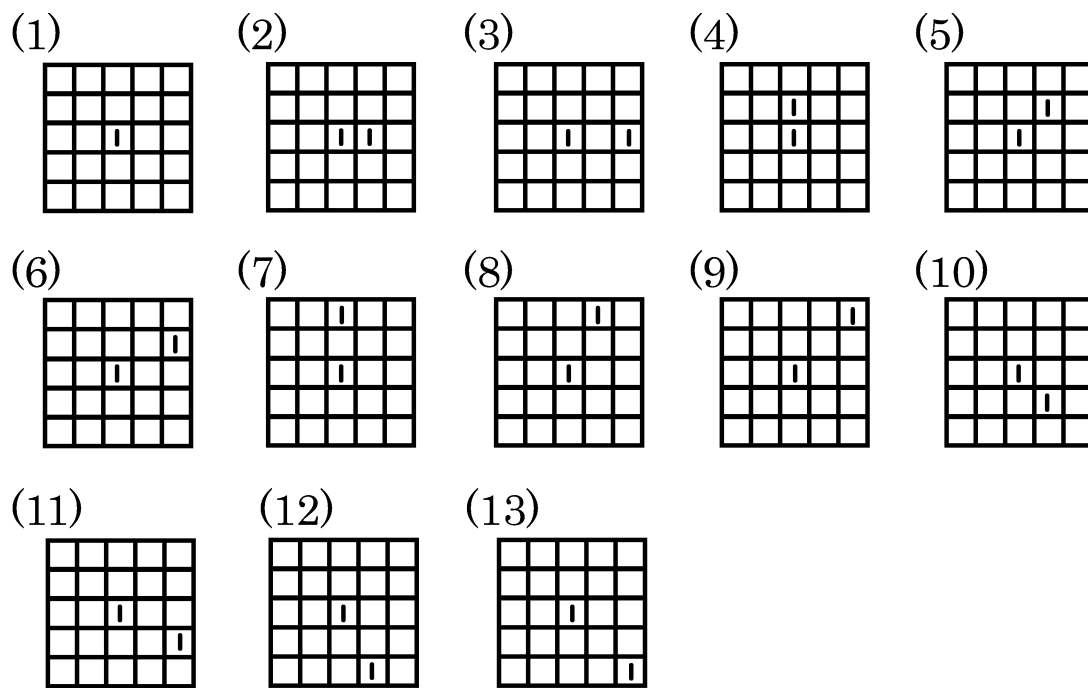


図 4.22: 局所領域の変位パターン

局所自己相関に基づく特徴は，近傍の $f(r)$ の積をピアノロール全体に対して足し合わせて得られた値である．本研究では $f(r)$ が1または0なので，その値はピアノロール上で図4.22の各パターンがそれぞれ何回出現したかを数えた値と等価になる．このような特徴は非常に局所的な特徴である．そこで，局所領域の1マスに入る範囲を音高方向で6種類（1, 2, 4, 6, 8, 12[半音]），伸縮後の時間方向で6種類（1, 2, 3, 4, 5, 6[伸縮後の時間]）に変化させ，詳細な情報からおおまかな情報までを抽出することを考えた．したがって，合計468（ $= 13 \times 6 \times 6$ ）個の特徴を抽出した¹⁰．

¹⁰このような階層的な特徴の抽出は，文献[栗田93]でも行われており，顔の輪郭の認識に有効であることが確認されている．

4.4.3 正準相関分析

MDA-OR を用いて意図空間上に配置した 100 個のフレーズに対して、意図ベクトルと局所自己相関特徴との正準相関分析を行った。468 個の特徴の中から、一番相関の高い特徴を 1 つずつ選択しながら正準相関分析するという操作を繰り返すことにより重要な特徴量を選択していったところ、50 個の特徴量を選択した時点で、正準相関の値が 0.95 を超え充分高くなったため、そこで終了とした。得られた相関を用いて、新たなフレーズの意図ベクトルが算出可能であることを確認した。

4.4.4 出力演奏の生成

意図空間上に配置したフレーズから、仮想演奏者のそのときの意図ベクトルに近いフレーズを次々と接続し出力演奏を生成する。その際、フレーズの接続部で違和感がないように考慮して接続する必要がある。そこで、フレーズを切り出した時にフレーズを囲むボロノイ線図には、もとの演奏記録における前後のフレーズのコンテキストがある程度残っていると考え、フレーズを生成する際に、そのフレーズを囲むボロノイ線図が交差したり、大きく隙間が開いたりしないようにすることを考える。具体的には、前後 2 つのフレーズを囲むボロノイ線図の重なりや隙間が、ある一定以上にならないように閾値を設定し、閾値を越えたフレーズは選択しないようにした。図 4.23 の (a) は、フレーズを切り取る前のボロノイ線図を表している。一方、(b) は、適切なフレーズを接続した例である。2 つのフレーズを囲むボロノイ線図の隙間や重なりが少なく、両者は滑らかに接続している。各フレーズは、12 小節の 1 コーラス中で、演奏記録からそのフレーズを切りだした位置と同じ位置のみで使用することにする。このようにすることで、コード進行の制約を満たした演奏が生成できる。

仮想演奏者が、演奏を生成する際に意図ベクトルに対応するフレーズが、フレーズデータベース上にない場合には、標準的なフレーズを多数収めてあるフレーズデータベースからフレーズを検索し演奏を生成する。このようにすることで、フレーズを模倣しようとする演奏者のフレーズの数が少ない場合でも、その演奏者のフレーズと標準的なフレーズの両方を用いて演奏を生成することができる。

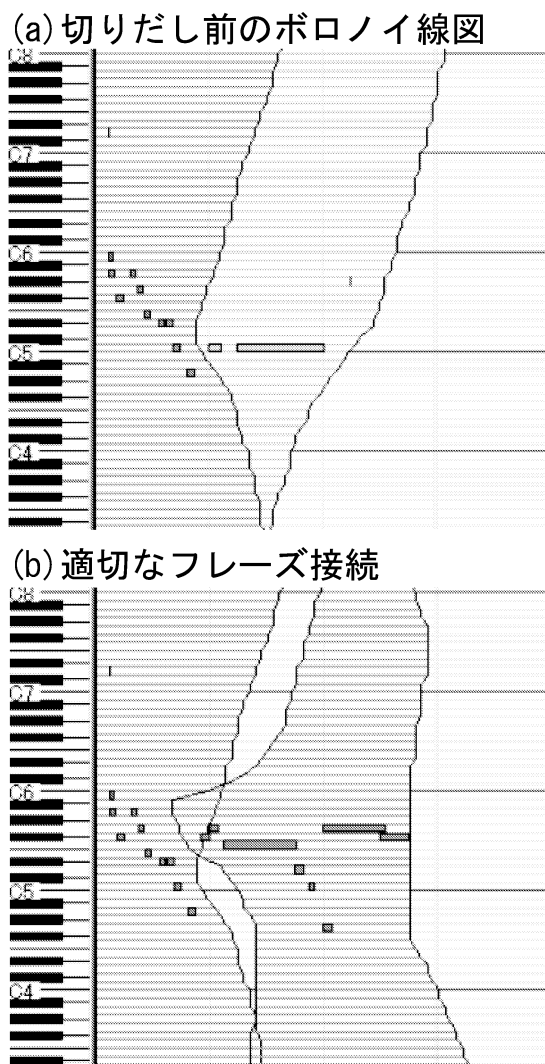


図 4.23: フレーズの接続

4.5 まとめ

本章では、ジャムセッションの演奏記録から演奏者固有のフレーズを獲得し、フレーズデータベースを生成して再利用するための手法として確率モデルに基づくクオンタイズ手法、ボロノイ線図を用いたフレーズの分割手法、MDA-OR と正準相関分析を用いたフレーズの再利用法について述べた。

4.2 節では、テンポ一定の伴奏に合わせて弾かれたゆらぎを含む即興演奏の発音時刻から、演奏者が元々弾こうとした正規化された発音時刻を推定する手法を提案した。そして、発音時刻の遷移とゆらぎを隠れマルコフモデルを用いてモデル化する

ことにより、和音を含む演奏を確率モデルで表すことを可能とした。正解データを用いた学習をすることにより、1つのHMMによるモデルが市販のシーケンスソフトウェアのクオンタイズを超える性能を持つことを示した。また、演奏の各拍に表れるパターンが16分系か8分3連系のいずれかしか含まない場合には、そのパターンに対応した4種類のHMMを用意することにより、性能が向上することを示した。ここでは、4種類のHMMによるモデルしか扱っていなかったが、今後32分音符などを含む曲のクオンタイズを考えた場合、HMMの数を増やしていく可能性について検討しなくてはならない。さらに、Baum-Welchアルゴリズムとヘルドアルト補間法により、HMMのモデルパラメータを教師なし推定する手法を提案し、教師データが与えられない場合でも高い性能を示すことが確認できた。提案手法によるクオンタイズが、発音時刻のゆらぎの大きさや演奏の傾向によらず安定して高い性能を示したことは、クオンタイズへの学習の導入が有効であることを表わしている。本研究で扱ったのは、テンポが一定で伴奏の拍位置が明らかな場合での演奏のクオンタイズであるが、可変テンポの場合への対応可能性についても検討していく。ここでは、学習によって得られたモデルパラメータを、実演奏データのクオンタイズに使ったが、5章では、各演奏者の出力確率分布を用いて、楽譜上に量子化された演奏への人間的なノリの付加（ヒューマナイズ）を試みる。

4.3節では、ポロノイ線図に基づくグルーピング手法を提案し、提案手法によるグルーピング結果がGTTMに基づき人手でグルーピングした結果に近くなることを確認した。これは、音高と時間のグルーピングルール競合という問題に対して、スケーリング比（音高/時間）という数値的（距離）尺度を明示的に与えたことが有効であったためだと考えられる。また、実際にジャムセッションの演奏記録をポロノイ線図に基づきフレーズ分割した結果も、人手による分割と近い結果となることが確認できた。

4.4節では、MDA-OR、局所自己相関関数、正準相関分析を用いて、分割されたフレーズを意図空間上に配置し再利用する手法について述べた。

第5章 発音時刻ゆらぎの模倣

本章では、4.2節でジャムセッションの演奏記録から学習した、発音時刻の順モデルを用いて、演奏に発音時刻ゆらぎの付加を試みる。模倣したい演奏者の演奏記録から学習した確率モデルを用いて、ゆらぎを付加することによって、その演奏者の発音時刻の特徴を反映した演奏の生成が期待できる。

5.1 はじめに

ジャムセッションなどで、演奏者がテンポ一定の伴奏に合わせて演奏した場合でも、演奏動作の微妙な差や演奏の表情づけの差により、発音時刻にゆらぎが生じる。ここでは、そのようなゆらぎの個性を模倣し、仮想演奏者の演奏に反映させことを考える。

発音時刻ゆらぎなど演奏の表情に関する研究としては、演奏レンダリングシステムの研究がある。演奏レンダリングシステムには、ルールベースによるシステム [FFBS91, Fri91, 青野 97, 石川 02, Wid94, Wid95]，ルールとニューラルネットワークを併用したシステム [白川 95, 小田 95, BPV92, CPRV97]，事例ベースによるシステム [鈴木 00, AdMS97]，ユーザと対話可能なシステム [平田 01b] などがある。

ルールベースによるシステム

文献 [FFBS91, Fri91] では、数値化した演奏ルールを設計し、それを計算機上を実装して演奏の表情づけを行った。文献 [Wid94, Wid95] では、音楽理論 [LJ83, Nar77] に基づき階層的な楽曲構造を分析し、機械学習により楽曲構造とそれに対応した表情曲線の組をルールとして獲得し、そのルールを用いて、未知曲に表情づけを行うシステムを実現した。文献 [青野 97, 石川 02] では、重回帰分析により、演奏ルールを説明変数の結合係数として獲得するシステムを実現した。文献 [石川 02] では、さらに曲の前半でルールの学習を行い、

後半に獲得したルールを適応することで、システムの評価を行っていた。

ルールとニューラルネットを併用したシステム

文献 [白川 95, 小田 95] では、演奏のペロシティの変化と発音時刻の大局的なゆらぎをルールベースを用いて付加し、発音時刻の局所的なゆらぎをニューラルネットに付加するシステムを試みた。ニューラルネットには、あらかじめ演奏のゆらぎと前後数個分の楽譜情報との関係をバックプロパゲーションにより学習させていた。文献 [BPV92, CPRV97] では、ルールにより楽譜を分析し、その分析結果を入力し演奏を出力するようなニューラルネットを構築した。

事例ベースによるシステム

文献 [鈴木 00] では、データベースに多くの演奏事例を収め、演奏生成時には演奏状況をキーとして、検索し、いくつかの選択された事例の表情を類似度に応じて重みづけして付加するシステムを実現した。文献 [AdMS97] では、事例に基づく推論を用いて、演奏に表情を付加するシステムを実現した。

ユーザと対話可能なシステム

文献 [平田 01b] では、ユーザの指示による逸脱と、システムによる音楽的制約の充足を繰り返すことにより、ユーザの望む演奏を生成することを提案している。プロトタイプシステムでは、模倣すべき事例曲の提示によってユーザの指示が行われており、事例ベースによるシステムに分類される。

演奏レンダリングシステムには、この他にも多くの研究事例がある。最新の内容については、文献 [BHH⁺02] 等を参考にされたい。

本章では、テンポ一定でジャムセッションするという条件下において、発音時刻ゆらぎの付加を扱う。限られた長さのジャムセッションの演奏記録から演奏者の発音時刻ゆらぎの癖や特徴を得るためには、4.2節で述べたような確率モデルによるアプローチが有効である。確率モデルは、ヘルドアウト補間法をはじめとするデータの補間法が確立されており少量の学習データを有効に使用することができる。また、確率分布は分布の形状を視覚的に確認することも容易なので、未学習の曲に対して予期しない出力が出る心配も少ない。以下、5.2節では、発音時刻ゆらぎの順モデルについて説明し、5.3節では、逆関数法を用いて、発音時刻ゆらぎの順モデルの確率分布に基づいた乱数の発生について説明する。そして、5.4節では、実際にフレーズデータベースに収められているゆらぎのない演奏にゆらぎの付加を行う。

5.2 発音時刻ゆらぎの順モデル

4.2.1でクオンタイズに用いた発音時刻ゆらぎの順モデルを用いてゆらぎのない演奏への発音時刻ゆらぎの付加を実現する(図5.1)。

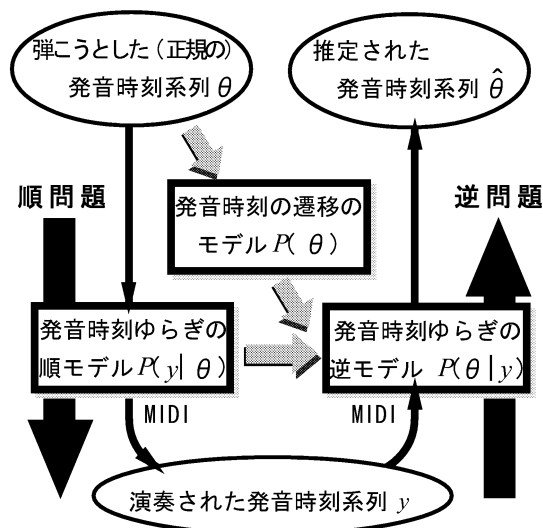


図 5.1: 発音時刻ゆらぎの順モデル

隠れマルコフモデルにより定式化し出力分布 $b_j(k)$ を学習した結果 (4.2.3 節), 発音時刻ゆらぎは, 図 5.2 のような, 正規分布に近い形をした分布で表される¹. つまり, テンポ一定の演奏にあわせて弾いた演奏の発音時刻ゆらぎの個性は, $b_j(k)$ の分布の中心と分散で表すことができる. そこでまず, $b_j(k)$ を正規分布で近似する. この近似は, 学習データの不足を補う意味と, 演奏者のミスになどより, 正規の発音時刻よりも大きく外れた音を取り除き, データを純化する意味がある. 近似後の正規分布の分散を $\frac{3}{4}$, 中心を μ とする. ただし μ は, $0 \leq \mu < 480$ の実数とする.

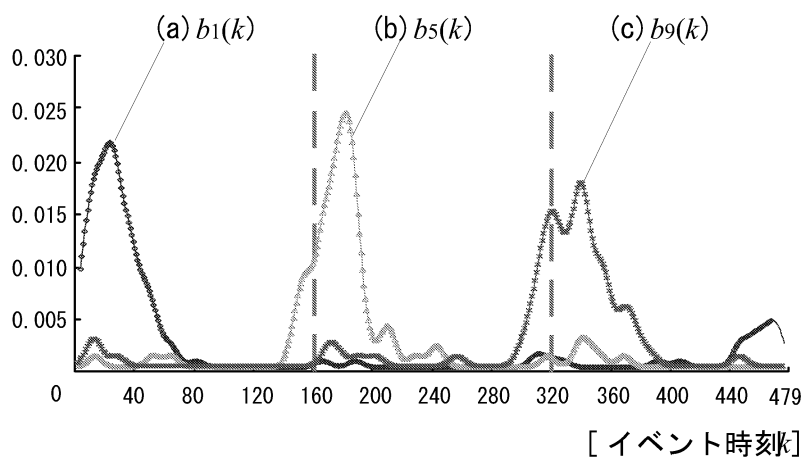


図 5.2: 実演奏データから求めた $b_1(k)$; $b_5(k)$; $b_9(k)$

5.3 逆関数法による確率分布乱数

逆関数法を用いて, 隠れマルコフモデルの出力分布 $b_j(k)$ に基づいた乱数を発生させる. r を区間 $[0; 1]$ の一様乱数, $\Phi^{-1}(\cdot)$ を, 標準正規分布の累積分布関数 (図 5.3) の逆関数とした場合, $b_j(k)$ を近似した正規分布に従う乱数 x は次式で表せる².

$$x = \Phi^{-1}(r) + \mu \quad (5.1)$$

¹図 5.2 は, 全体的に正規の発音時刻 (図の Y 軸および点線の位置) より遅れて発音している, すなわち, もたつた演奏をしている例である.

²ここで乱数を導入したのは, 発音時刻がどの値をとっても妥当であるためである. 乱数を使っているために, 再現性は低下しているが, この手法で出力した演奏の演奏記録から得られる確率分布と, ゆらぎに用いた確率分布は, 演奏の長さが長くなるにつれて一致していく.

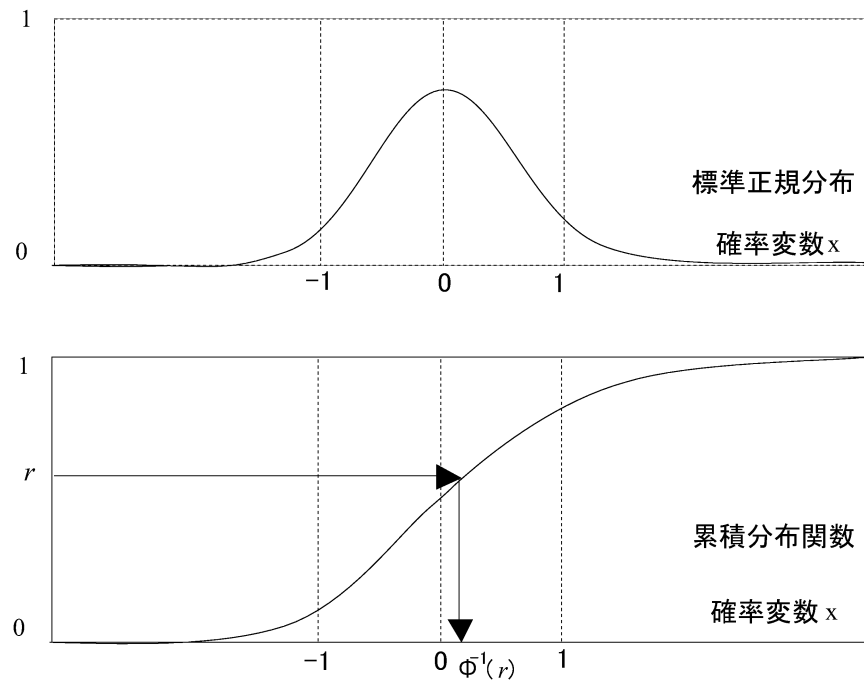


図 5.3: 標準正規分布とその累積分布関数

5.4 発音時刻ゆらぎの付加

発生させた乱数 x を用いて、発音時刻にゆらぎを付加する。このとき、和音の構成音のそれぞれに同じ確率分布でゆらぎを付加してしまうと、ゆらぎの付加の前の和音の構成音の発音順序と後の発音順序が異なってしまふ場合があり問題がある。なぜなら、ギターは弦が6本あり同じ構成音の和音でも、ダウンピッキング（上から下）で弾いた場合と、アップピッキング（下から上）で弾いた場合では、ニュアンスが異なるからである。図5.4は、3和音に同じ確率分布でゆらぎを付加した場合の例である。まず（a）は、ゆらぎを含む演奏記録で（b）は確率モデルに基づくクォンタイズによりゆらぎを取り除いた状態である。フレーズデータベースは（b）のように発音時刻を正規化した状態の演奏が収めている。このとき、和音の発音時刻は、すべての構成音で同じになるが、MIDI ファイル上では発音順序は保存されている。そして（c）は、演奏生成時に、和音のすべての構成音に、同じ確率分布でゆらぎを付加した結果である（a）と（c）では、和音の構成音の発音順序が異なっている。

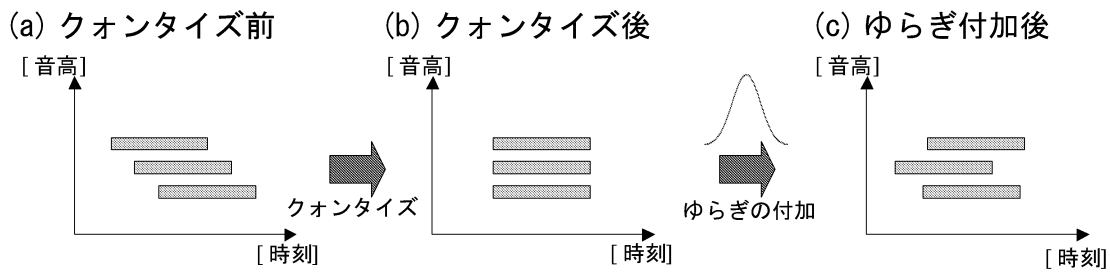


図 5.4: 和音へ同じ確率分布でゆらぎを付加した例

本研究では、ゆらぎを付加する前後で、和音の発音順序が変化しないように以下のような処理を行う。まず、 $b_j(k)$ を近似した正規分布を和音の構成音の数で、分布の面積をイベント時刻方向に等分割した分布を作る (図 5.5)。そして、それを面積が1になるように正規化した分布に基づく乱数でゆらぎを付加する。具体的には、以下の式を用いてゆらぎを付加した。ただし、 n は和音の構成音の数、 m はそのうち何番目の音であるかを表す。

$$x = \epsilon \odot i^{1-\frac{m}{n}} + 1 \quad (5.2)$$

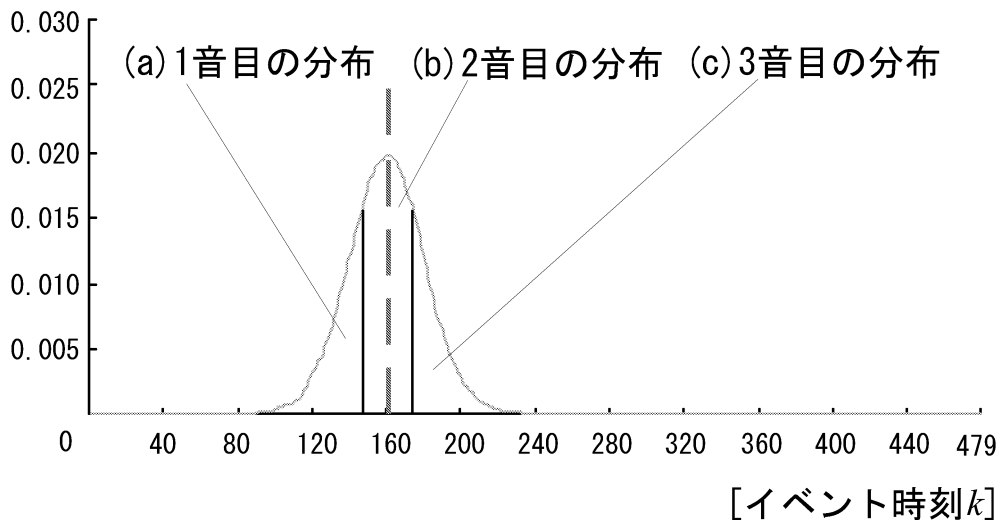
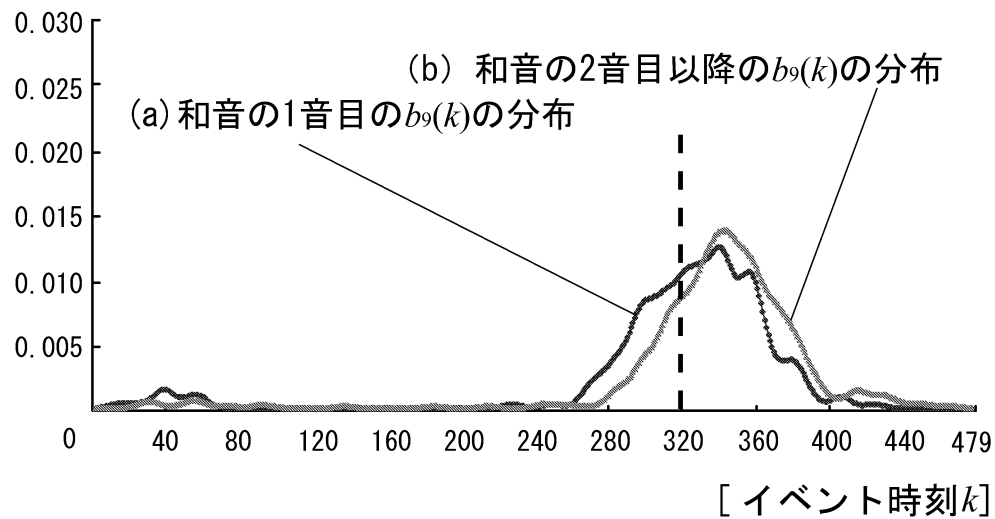


図 5.5: 3和音の場合における分布の分割

図 5.6: 和音の1音目と2音目以降の $b_9(k)$

5.5 ゆらぎの付加に関する考察

4.2節で提案した隠れマルコフモデルは、もともとクオンタイズのために設計したモデルであったため、そのモデルを発音時刻ゆらぎの付加に利用するためにはいくつか問題が生じた。最大の問題は、発音時刻でモデル化しているために、和音の構成音がすべて同じ状態となることである。このモデル化は、演奏をクオンタイズする際には利点であったが、ゆらぎを付加する際には逆に欠点となり、ゆらぎを付加する前後で、和音の構成音の順序が異なってしまう場合が生じる。そこで5.4節では、ゆらぎを付加する前後で、和音の構成音の順序が保存されるように、和音の場合にはその構成音の数で等分割した分布を用いてゆらぎを付加した。このような単純な処理で、和音の発音順序を維持しながらゆらぎの付加を行ったのは、4.2節で発音時刻ゆらぎの特徴を学習した隠れマルコフモデルを有効に利用したいと考えたためである。

しかし、実際の和音の発音時刻分布は図5.5のように等分割されているわけではない。図5.6は、4.2節の演奏B2に含まれている和音の第1音目の $b_9(k)$ の分布と、2音目以降の $b_9(k)$ の分布を比較したものである。その結果、第1音目の $b_9(k)$ の分布のほうがわずかに左にずれた分布となったが、両者はほとんどの部分で重なっていた。一方、図5.7は、同じ演奏の和音の構成音の発音時刻間隔IOI (Inter-Onset-Interval) の分布である。分布は、正規分布の右側半分に近い形状となった。

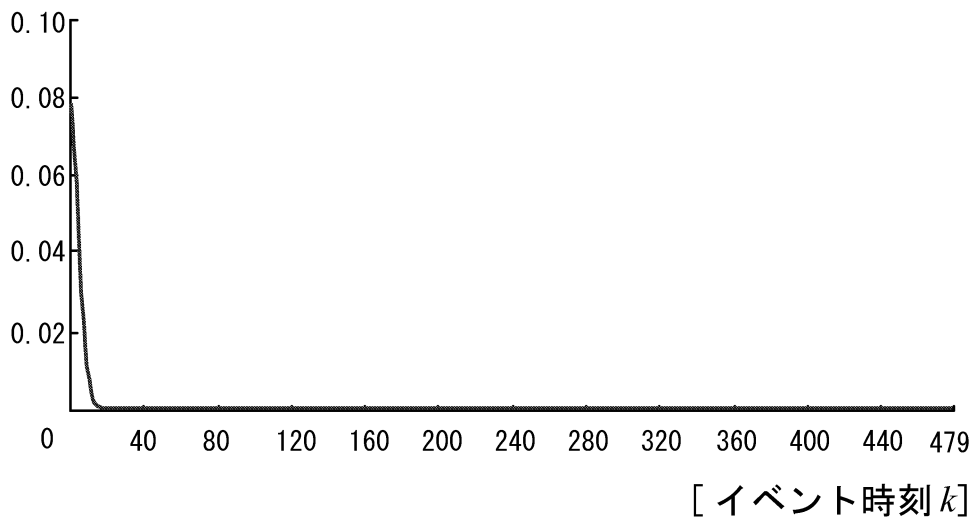


図 5.7: 和音の発音時刻間隔 IOI の分布

今後、和音にゆらぎを付加する過程を詳細にモデル化していくためには、これらの結果を踏まえて、発音時刻ゆらぎの付加に特化した新たなモデルを構築していく必要がある。

5.6 まとめ

本章では、ある演奏者の演奏記録から、発音時刻ゆらぎの個性を学習した隠れマルコフモデルの出力分布 $b_j(k)$ を使って、フレーズデータベースに収められているゆらぎのない演奏の発音時刻にゆらぎの付加を試みた。和音の場合には、和音の構成音の数で等分割した分布を用いてゆらぎを付加した。

第6章 評価と考察

本章では、以上述べてきたシステムの評価と考察を行う。ジャムセッションシステムには、絶対的な評価の枠組みはまだ確立されていない。本研究では、定量的な評価と心理実験による評価の両方でシステムの評価を行う。

6.1 はじめに

ジャムセッションシステムなどの音楽システムの評価の方法については、従来から様々な議論がなされてきている [平田 97, 平田 01a, 片寄 02a] が、本研究では、定量的な評価と心理実験による評価の両方を行うのが妥当であると考えている。システムの出力に対して、その理想的な出力（正解データ）が用意できる場合には、定量的な評価が有効である。一方、仮想演奏者が演奏者を模倣できていることを示す上では、心理実験による評価が有効である。心理実験の方法としては、チューリングテスト [RN95] のような手法が考えられる。チューリングテストでは、相手が人間であるか機械であるかを判断したが、ここでは、どの演奏者の個性を模倣した仮想演奏者なのかを判断する。

6.2 定量的な評価

本ジャムセッションシステムは、いくつもの要素技術を用いて構築しており、それぞれの要素技術の性能について、できる限り定量的な評価を行ってきた。本研究で行った定量的評価について、以下にまとめる。

2 演奏者の振る舞いのモデルの評価

3章では、人間の演奏者 C と仮想演奏者 A, B（仮想演奏者 B は人間 C のモデルをもつ）でジャムセッションを行い、仮想演奏者 B が人間 C の振る舞い

を模倣できているかを評価した。その結果、12 コーラス分の演奏の各時刻における人間 C の意図ベクトルと、仮想演奏者の意図ベクトルの距離を合計し、その平均値を比較したところ、人間 C と仮想演奏者では、平均 9.33 と離れていたが、人間 C と仮想演奏者 B では、平均 3.56 と近かった。これは、人間 C の意図ベクトルと仮想演奏者 B の意図ベクトルが似通った動きをしていることを示す。このことから、仮想演奏者が人間の振る舞いを模倣できていることが定量的に確認できた。

3章では、その他、3.3.1 の印象算出部で、システムが示す印象ベクトルと被験者の評定値とを比較するなど、システムの構築過程でも定量的な評価を行った。

2 確率モデルに基づくクオンタイズの評価

4.2 節では、確率モデルに基づくクオンタイズの性能を評価した。まず、市販のシーケンスソフトウェアに搭載されている閾値処理による機械クオンタイズと教師つき学習した HMM クオンタイズを比較した結果、HMM クオンタイズが機械クオンタイズを超える性能であることを確認した。次に、1 拍内の音符が 8 分連系か 16 分系のいずれかであるという制約をつけた、制約つき機械クオンタイズと教師つき学習した HMM クオンタイズを比較した結果、実演奏データでは、HMM クオンタイズが制約つき機械クオンタイズを超える性能であることを確認した。1 つの HMM によるモデルと 4 つの HMM によるモデルを比較した結果、教師つき学習では 4 つの HMM によるモデルが性能が高く、教師なし学習では 1 つの HMM によるモデルが性能が高かった。教師つき学習と教師なし学習を比較した結果、教師なしは教師つきより若干低い性能となったが、すべての演奏で 7 割以上の一致率を示し、教師なしが高い性能を示すことが確認できた。

2 ボロノイ線図に基づくグルーピング・フレーズ分割の評価

4.3 節 4.3.3 では、ボロノイ線図に基づくグルーピング結果が GTTM に基づき手作業でグルーピングした結果とどれだけ一致しているかを一致率で評価した。その結果、3 種類のスケーリング比すべてで、72.8 % 以上という高い一致率であった。

4.3.4 では、ボロノイ線図に基づくフレーズ分割手法が、人間の手作業による

フレーズ分割結果とどれだけ一致しているかを一致率で評価した。その結果、すべての演奏者で 70.6 % 以上という高い一致率であった。

以上のような定量的な評価は、多くの場合システムの出力と正解データ（理想的な出力）とを比較することによって得られる。現状では、クオンタイズやグルーピング、フレーズ分割などの正解データを収めたデータベースはなく、本研究では、数曲の正解データを手作業で作成し評価を行った。今後、これらの評価のベンチマークとして利用できる共通のデータベースの整備が望まれる¹。

6.3 心理実験による評価

被験者が演奏者の個性が模倣できているかを評価する聴取実験を行う。被験者は、3人でいずれも演奏歴 10 年以上のギタリスト A,B,C である。演奏者 A,B,C はお互いに他の 2 人と 1 年以上同じバンドでの演奏経験を持つ者を選んだ。これは、誰の個性を模倣しているか判断する際に、相手の演奏をよく把握しているほうが好ましいからである。また、3人とまったく面識のない演奏者 D,E,F を用意した。実験は以下の手順で行った。

1. 演奏者 A,B,C,D,E,F について、人間の演奏者 1 人と仮想演奏者 2 人という形式でジャムセッションを実施しその演奏記録を得た。その際、演奏中に 6 人の演奏者がお互いに接触しないようにした。演奏時間は 5 分（10 コーラス）、テンポは 120 である。
2. 演奏記録から、それぞれの演奏者の振る舞いの個性、演奏者固有のフレーズ、発音時刻ゆらぎの個性を獲得し、それらを使って、仮想演奏者を用意した。評価に用いる仮想演奏者は、演奏者 A,B,C を模倣した仮想演奏者および、3 種類の個性を A,B,C,D から別々に選んで作った仮想演奏者（たとえば、演奏者の振る舞いの個性と演奏者固有のフレーズを A から模倣し、発音時刻ゆらぎの個性を B から模倣した仮想演奏者など）合計 39 種類の仮想演奏者を用意した（表 6.1）。

¹研究者が研究目的に使用する上で、共通利用の自由、学術利用の自由が確保された研究用音楽データベースとしては、ジャンル別に分類された曲とその曲の MIDI ファイル、トラックダウン前の音響信号などを収めた RWC 研究用音楽データベース [後藤 01, 後藤 02a, 後藤 02b] がある。

3. 仮想演奏者の生成した演奏が、A,B,C,Dのうち誰を模倣した仮想演奏者であるかを、演奏者A,B,Cに評定させた。A,B,Cのいずれでもないと思った場合にはDだと評定するように指示した。演奏は、評定対象の仮想演奏者（39種類）と、演奏者E,Fの個性を模倣した2人の仮想演奏者とのジャムセッションの演奏記録で、長さは2分である。3人の演奏のうち、どれが評定対象かわかるように、仮想演奏者の音色をそれぞれ異なるものに設定した。

6.4 実験結果・考察

表6.1に心理実験の結果を示す。3種類の個性を同じ演奏者から模倣している場合（仮想演奏者1,14,27）では、その演奏者を正しく評定できた割合は100%であった。3種類の個性のうち、振る舞いの個性以外の個性を同じ演奏者から模倣している場合（仮想演奏者2,5,8,13,15,18,21,26,28,33,34,39）では、その演奏者を正しく評定できた割合は、77.8%であった。3種類の個性のうち、演奏者固有のフレーズ以外の個性を同じ演奏者から模倣している場合（仮想演奏者3,6,9,12,16,19,22,25,29,32,35,38）では、その演奏者を正しく評定できた割合は、22.2%であった。3種類の個性のうち、発音時刻ゆらぎの個性以外を同じ演奏者から模倣している場合（仮想演奏者4,7,10,11,17,20,23,24,30,31,36,37）では、その演奏者を正しく評定できた割合は、86.1%であった。上記の結果を図6.1に表す。すべての個性

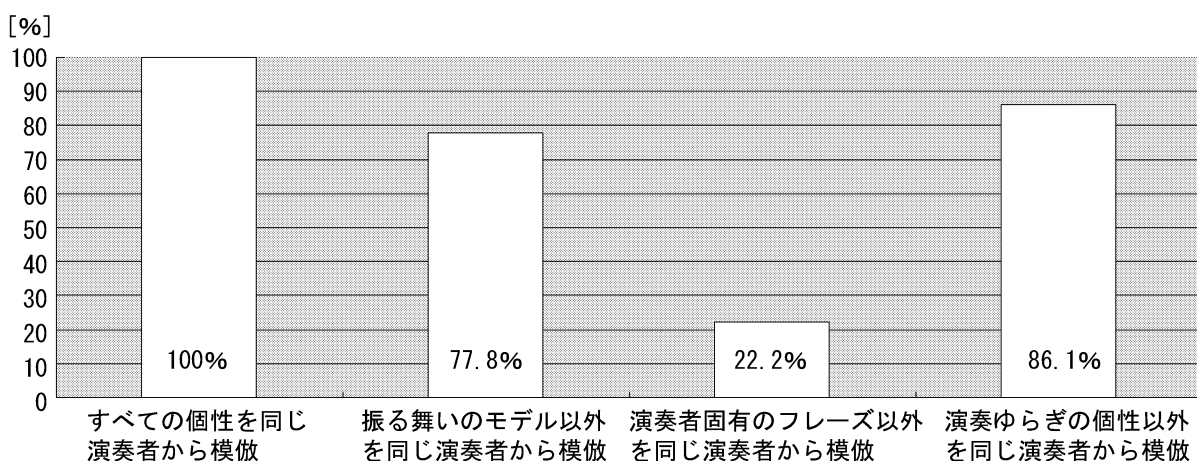


図 6.1: 評定結果と模倣した演奏者が一致した割合

を同じ演奏者から模倣している場合に比べ、3種類の個性のうち、どれか1つでも他の演奏者の個性を用いた場合には、正しく評定される割合が減っていた。これは、3つの個性のいずれもが演奏者の個性を模倣しており、仮想演奏者が、元の演奏者を模倣していると評定されるためには、3つの個性のすべてが必要であることを表している。

一方、3つの個性のうち、どれか1つを他の演奏者から模倣した場合どうしを比較すると、仮想演奏者の模倣元の演奏者を正しく評定するためには、演奏者固有のフレーズが最も重要で、次に演奏者の振る舞いの個性が重要で、それに比べて発音時刻ゆらぎの個性はそれほど重要ではないという結果になった。

ただし、これは被験者による聴取実験の結果であり、実際にジャムセッション中の演奏者にとっては、演奏者の振る舞いの個性も非常に重要な要素であると考えられる。また、発音時刻ゆらぎの個性がそれほど重要でないという結果になったことには、2つの理由がある。まず1つ目の理由は、発音時刻ゆらぎのモデルが、和音の場合では詳細なモデル化ができていないことである。2つ目の理由は、今回の実験ではテンポ一定でしかもテンポを120という高い値に設定していたことである。今後、和音の場合における詳細なモデル化を行い、低いテンポやテンポ可変の条件で実験を行った場合、発音時刻ゆらぎの個性が、より重要な要素となる可能性がある。

6.5 まとめ

本章では、システムの評価として、定量的な評価と心理実験による評価の両方を行った。まず、定量的な評価では、演奏者の振る舞いのモデルの評価、確率モデルに基づくクオンタイズの評価、ポロノイ線図に基づくグルーピング・フレーズ分割の評価を行った。そして、チューリングテストからヒントを得た心理実験を行い、仮想演奏者が演奏者の個性を模倣できているかを確認した。その結果、3つの個性のいずれもが演奏者の個性を模倣できていることを確認した。

表 6.1: 心理実験の結果

	仮想演奏者の個性			評定結果		
	振る舞いの個性	固有のフレーズ	ゆらぎの個性	演奏者 A	演奏者 B	演奏者 C
仮想演奏者 1	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>
仮想演奏者 2	B	A	A	A	A	A
仮想演奏者 3	A	B	A	B	D	A
仮想演奏者 4	A	A	B	A	A	A
仮想演奏者 5	C	A	A	A	D	A
仮想演奏者 6	A	C	A	C	D	C
仮想演奏者 7	A	A	C	A	A	A
仮想演奏者 8	D	A	A	A	D	A
仮想演奏者 9	A	D	A	D	D	D
仮想演奏者 10	A	A	D	A	A	A
仮想演奏者 11	D	D	A	A	A	D
仮想演奏者 12	D	A	D	A	D	A
仮想演奏者 13	A	D	D	D	D	D
仮想演奏者 14	<u>B</u>	<u>B</u>	<u>B</u>	<u>B</u>	<u>B</u>	<u>B</u>
仮想演奏者 15	A	B	B	B	B	B
仮想演奏者 16	B	A	B	A	D	B
仮想演奏者 17	B	B	A	B	B	B
仮想演奏者 18	C	B	B	B	D	B
仮想演奏者 19	B	C	B	C	D	C
仮想演奏者 20	B	B	C	B	D	B
仮想演奏者 21	D	B	B	B	D	B
仮想演奏者 22	B	D	B	D	D	B
仮想演奏者 23	B	B	D	B	B	B
仮想演奏者 24	D	D	B	D	B	D
仮想演奏者 25	D	B	D	B	D	B
仮想演奏者 26	B	D	D	D	D	D
仮想演奏者 27	<u>C</u>	<u>C</u>	<u>C</u>	<u>C</u>	<u>C</u>	<u>C</u>
仮想演奏者 28	A	C	C	C	D	C
仮想演奏者 29	C	A	C	A	D	A
仮想演奏者 30	C	C	A	C	C	C
仮想演奏者 31	C	C	B	C	C	C
仮想演奏者 32	C	B	C	B	D	C
仮想演奏者 33	B	C	C	C	D	C
仮想演奏者 34	D	C	C	C	D	C
仮想演奏者 35	C	D	C	D	D	C
仮想演奏者 36	C	C	D	C	C	C
仮想演奏者 37	D	D	C	D	C	D
仮想演奏者 38	D	C	D	C	D	C
仮想演奏者 39	C	D	D	D	D	C

第7章 結論

本章では、本論文で述べてきた研究成果についてまとめ、今後の課題と、ジャムセッションシステムの研究の行方について考察する。

本論文では、実在する人間の個性を模倣した仮想演奏者と人間の演奏者とがジャムセッションできるシステムについて述べた。具体的には、演奏者の振る舞いの個性、演奏者固有のフレーズの個性、発音時刻ゆらぎの個性をそれぞれ獲得し、仮想演奏者にその個性を反映させた。

第1章では、本研究の背景、関連研究についてまとめ、本研究の位置づけを明らかにした。本研究は、ジャムセッションシステムという題材を通して、人間の個性の工学的な扱いを確立することを目的としている。ジャムセッションにおける演奏者の個性は、その演奏者が参加したセッションの演奏記録に反映されており、演奏記録を統計的に学習することにより演奏者モデルの獲得を実現した。ジャムセッションシステムを構成する各要素は、音楽情報科学の研究分野において重要な技術であり、今後、他の音楽システムへの利用も期待できる。

第2章では、ジャムセッションにおける演奏者の個性として演奏者の振る舞いの個性、演奏者固有のフレーズの個性、発音時刻ゆらぎの個性について述べ、システムの全体像について説明した。そして、演奏者の個性を獲得してシステムを実現する上での問題点についてまとめた。システムは、演奏者の個性を収めた3種類の個性データベースと、その個性を用いて動作する仮想演奏者からなる。演奏者の振る舞いの個性は、模倣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獲得した。演奏者固有のフレーズは、模倣しようとしている演奏者の演奏を、再利用可能な形でフレーズに分割することにより獲得した。発音時刻ゆらぎの個性は、発音時刻の特徴・癖を演奏記録から確率モ

デルとして学習することにより獲得した。

第3章では、演奏者の振る舞いを学習する手法について述べ、実際に演奏記録から振る舞いのモデルの学習を実現した。そして、獲得した振る舞いのモデルに基づき仮想演奏者が動作することを確認した。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ができる。しかし、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、まず入力演奏から印象ベクトルを算出し、次に出力演奏から意図ベクトルを算出し、最後に演奏者の振る舞いのモデルを学習するという3段階の処理を考案した。

第4章では、演奏者固有のフレーズを模倣する手法として、発音時刻のクオンタイズ、演奏のフレーズ分割、演奏の再利用法について説明した。発音時刻のクオンタイズでは、確率モデルを用いて、ジャムセッションの演奏のゆらぎを含む発音時刻から、演奏者が元々弾こうとした正規化された発音時刻の推定を実現した。演奏のフレーズ分割では、ポロノイ線図を用いて、和音を含むジャムセッションの演奏を再利用可能なフレーズへの分割を実現した。演奏の再利用法では、獲得した多数のフレーズをMDA-OR、局所自己相関関数、正準相関分析を使って意図空間上に配置し、配置したフレーズを使って演奏の生成を実現した。

第5章では、発音時刻ゆらぎを模倣する手法について述べ、ゆらぎのない演奏に実際にゆらぎの付加を行った。クオンタイズに用いた確率分布は、発音時刻の特徴や癖を獲得している。そこで、逆関数法を用いて確率分布に基づく乱数を発生させることにより、模倣したい演奏者の発音時刻の特徴を反映した演奏の生成を実現した。ゆらぎを付加する音が和音の場合、和音の構成音の数で、確率分布を等分割した分布を用いることで、和音の構成音の発音順序を変化させずにゆらぎを付加することができた。

第6章では、構築したシステムの評価を行った。ジャムセッションシステムには、まだ絶対的な評価の枠組みが確立されていない。そこで本研究では、定量的な評価と心理実験による評価の両方でシステムを評価した。チューリングテストからヒントを得た心理実験を行い、仮想演奏者が演奏者の個性を模倣できているかを確認し

た結果，演奏者の振る舞いの個性，演奏者固有のフレーズ，発音時刻ゆらぎの個性のいずれもが模倣できていると判断できた．

今後の課題として，第一に，セッションに参加する人数を可変とする拡張があげられる．本研究で実装したシステムは，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のものであったが，本研究で扱ったセッションモデル（2章の図2.3）自体は，任意の人数に拡張できるものである．実際，2人の演奏者にシステム上でセッションを行わせたところ，演奏者の振る舞いのモデルが獲得でき，2人でのセッションにも適用できる可能性があることがわかった．今後，4人以上のセッションの場合も試みていきたい．

第二の課題として，学習アルゴリズムの改良があげられる．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を用いて学習したため，セッション全体の展開や流れを把握した演奏は実現できていない．ジャズの曲などのように，コーラス数が決まっていたり，テーマを決めて演奏するような場合には，各演奏者がどのコーラスでどんな演奏をしたいかなど，今回扱ったものより，もっと大局的な演奏意図を考慮していく必要がある．今後，そのような大局的な演奏意図を扱うための手法について検討するとともに，大局的に見た時系列の変化も考慮にいれた学習アルゴリズムについて検討していく予定である．

第三の課題として，ジャムセッションをする上での制約を減らしていくことが考えられる．本システムでは，テンポ一定で，コード進行一定の条件下でのジャムセッションを行ったが，今後，テンポやコード進行が自由となるようなシステムを実現していく．

今後，システムおよび，獲得した演奏者の個性をネットワーク上で配信し，多くの人に利用してもらうことを考えている．すなわち，一般公開をすることによりユーザを増やし，使用したユーザの個性を次々にデータベースに追加していくことで，様々な演奏者の個性を収集することを考えている．また，将来的には，著名なアーティストの個性をパッケージ化することも検討していきたい．

現在ジャムセッションシステムの研究は，各研究グループごとに個別に行われており，他のシステムの仮想演奏者と，自分の開発したシステムとを比較・検討することが困難な状況にある．今後，システムの公開を進めていくことによって，別々の

システムの仮想演奏者でセッションを行い、それを評価するなど様々な試みを行っていきたい。

本研究により、ジャムセッションにおける演奏者の個性を模倣した仮想演奏者が実現可能であることが示された。今後、鑑賞、演奏、作曲・編曲という、人間の音楽的活動を統合的に行うことのできる仮想演奏者を計算機上で実現していきたい。

謝辞

本論文は、筆者が筑波大学大学院 工学研究科 に在籍し、連携大学院生として産業技術総合研究所に於いて行ってきた研究をまとめたものです。

本論文をまとめるに際し、筑波大学連携大学院の2年次以来4年間にわたり指導教官として御指導と御鞭撻を賜りました、大津 展之 教授（産業技術総合研究所，東京大学）に心から感謝の意を表します。また、同じく連携大学院の初年度に指導教官として、本研究を開始する機会と御指導を賜りました、築根 秀男 教授（産業技術総合研究所）に深く感謝の意を表します。本論文の審査にあたって、貴重な御助言と御指導を頂いた、筑波大学 システム情報工学研究科の板橋 秀一 教授，寅市 和男 教授，北脇 信彦 教授，片岸 一起 助教授に深く感謝いたします。

本研究は、産業技術総合研究所の多くの方々のご支援のおかげで行うことができました。連携大学院生として快く受け入れて頂いた、大蒔 和仁 情報処理研究部門長，浅野 太 メディアインタラクショングループ長に深く感謝申し上げます。麻生 英樹 主任研究員には、4.2 節で用いた隠れマルコフモデルなど、統計的な学習手法について懇切なる御指導，御助言を頂きました。後藤 真孝 博士には、全演奏者が対等なセッションを実現するためのモデルや、システムのリアルタイムでの実装，出力演奏の生成などについて，本ジャムセッションシステムを構築する上で必要不可欠な御助言，御指導を頂きました。また，後藤 氏には，国際会議や国内での発表の準備など，4年間にわたり多大な協力を頂きました。以上の方々に深く感謝いたします。

是永 定美 博士（元日本大学，元電子技術総合研究所），加藤 俊一 教授（現中央大学，元電子技術総合研究所）には大学4年次に研究所で研究を開始する機会を与えて頂きました。ここに深く感謝申し上げます。

NTTコミュニケーション基礎科学研究所には、短期間滞在する機会を与えて頂きました。村瀬 洋 メディア情報研究部長、小坂 直敏 メディア表現研究グループ長に深謝いたします。4.3 節のポロノイ線図によるグルーピングは、平田 圭二 博士(メディア情報研究部 特別研究員) との共同研究の成果を発展させたものです。その際には、議論や実装を通じて、本研究を遂行する上でも有用な経験をさせて頂きました。ここに深謝いたします。

著者が研究を開始するきっかけとなったのは、大学1年次に早稲田大学ファジィ研究会(山下研究室)に参加したことでした。山下 元 教授(早稲田大学)には、暖かい御理解と御援助を頂き、謹んで御礼申し上げます。同研究会のメンバーには、研究を遂行する上で有益な御示唆を頂きました。とりわけ稲井田 次郎 教授(日本大学)には、細やかな御指導を頂きました。ここに感謝の意を表します。

情報処理学会 音楽情報科学研究会では貴重な御意見を頂きました。研究会のメンバーおよび主査に感謝の意を表します。

演奏者として実験に協力して頂いた、斎田 康広 氏、千田 真一 氏、橋本 大輔 氏に感謝いたします。

本研究の遂行に際しては、この他多くの方々から、御指導、御助言、御援助を頂きました。ここに記して深謝いたします。

最後に、妻としてまた研究者として、著者の研究生活を支えて頂いた、昇姫に心から感謝いたします。

参考文献

- [AD90] Paul E. Allen and Roger B. Dannenberg. Tracking musical beats in real time. Proceedings of the 1990 International Computer Music Conference, pp. 140{143, September 1990.
- [AdMS97] Josep Lluís Arcos, Ramon López de Mantaras, and Xavier Serra. Saxex: A case-based reasoning system for generating expressive musical performances. Proceedings of the 1997 International Computer Music Conference, pp. 329{336, September 1997.
- [AKI95] Yushi Aono, Haruhiro Katayose, and Seiji Inokuchi. An improvisational accompaniment system observing performer's musical gesture. Proceedings of the 1995 International Computer Music Conference, pp. 106{107, September 1995.
- [AW97] Christina Anagnostopoulou and Gert Westermann. Classification in music: A computational model for paradigmatic analysis. Proceedings of the 1997 International Conference of Music Perception and Cognition, pp. 125{128, September 1997.
- [BHH⁺02] R. Bresin, P. Desain M. Hashida, R. Hiraga, O. Ishikawa, H. Katayose, Y. Miyagawa, K. Nishimoto, K. Noike, K. Ohgushi, C. Oshima, M. Sadakata, T. Shirosaki, T. Suzuki, Y. Takeuchi, A. Tobudic, and Y. Uwabu G. Widmer. Proceedings of ICAD2002 Rencon Workshop, July 2002.
- [BPV92] Roberto Bresin, Giovanni De Poli, and Alvisè Vidolin. Symbolic and sub-symbolic rules system for real time score performance. Proceedings of the

- 1992 International Computer Music Conference, pp. 211{214, October 1992.
- [Cam01] Emiliios Cambouropoulos. The local boundary detection model (LBDM) and its application in the study of expressive timing. Proceedings of the 2001 International Computer Music Conference, pp. 290{293, September 2001.
- [CCG91] S. Chen, C. F. N. Cowan, and P. M. Grant. Orthogonal least squares learning algorithm for radial basis function networks. IEEE Transactions Neural Networks, Vol. 2, No. 2, pp. 302{309, March 1991.
- [Cop92] David Cope. Computer modeling of musical intelligence in EMI. Computer Music Journal, Vol. 16, No. 2, pp. 69{83, Summer 1992.
- [CPRV97] Sergio Canazza, Giovanni De Poli, Antonio Roda, and Alvisè Vidolin. Analysis by synthesis of the expressive intentions in musical performance. Proceedings of the 1997 International Computer Music Conference, pp. 113{120, September 1997.
- [Del87] Irene Deliège. Grouping conditions in listening to music: An approach to Lerdahl and Jackendo®'s grouping preference rules. Music Perception, Vol. 4, No. 4, pp. 325{360, 1987.
- [DH89] Peter Desain and Henkjan Honing. The quantization of musical time: A connectionist approach. Computer Music Journal, Vol. 13, No. 3, pp. 56{66, Fall 1989.
- [DMR87] Roger B. Dannenberg and Bernard Mont-Reynaud. Following an improvisation in real time. Proceedings of the 1987 International Computer Music Conference, pp. 241{248, August 1987.
- [EM77] Bradley Efron and Carl Morris. Stein's paradox in statistics. Scientific American, Vol. 236, No. 5, pp. 119{127, March 1977.

- [FFBS91] Anders Friberg, Lars Fryden, Lars-Gunnar Bodin, and Johan Sundberg. Performance rules for computer-controlled contemporary keyboard music. *Computer Music Journal*, Vol. 15, No. 2, pp. 49{55, Summer 1991.
- [Fri91] Anders Friberg. Generative rules for music performance: A formal description of a rule system. *Computer Music Journal*, Vol. 15, No. 2, pp. 56{71, Summer 1991.
- [GHM⁺96] Masataka Goto, Isao Hidaka, Hideaki Matsumoto, Yosuke Kuroda, and Yoichi Muraoka. A jazz session system for interplay among all players | VirJa Session (virtual jazz session system) | . Proceedings of the 1996 International Computer Music Conference, pp. 346{349, August 1996.
- [GM95] Masataka Goto and Yoichi Muraoka. A real-time beat tracking system for audio signals. Proceedings of the 1995 International Computer Music Conference, pp. 171{174, September 1995.
- [Hal94] Peter Halász. Computer simulation of time-span reduction. Proceedings of the 1994 International Conference of Music Perception and Cognition, Vol. 4, No. 4, pp. 267{268, July 1994.
- [HHH⁺02] Rumi Hiraga, Mitsuyo Hashida, Keiji Hirata, Haruhiro Katayose, and Kenji Noike. RENCON: toward a new evaluation method for performance rendering system. Proceedings of the 2002 International Computer Music Conference, pp. 357{360, September 2002.
- [HM02] Keiji Hirata and Shu Matsuda. Interactive music summarization based on GTTM. Proceedings of the 3rd International Conference on, pp. 86{93, October 2002.
- [KW78] Joseph B. Kruskal and Myron Wish. *Multidimensional Scaling. Quantitative Applications in the Social Sciences*. Sage Publications, January 1978.
- [LJ83] Fred Lerdahl and Ray Jackendo®. *A Generative Theory of Tonal Music*. The MIT Press, 1983.

- [MR68] J. A. McLaughlin and J. Raviv. Nth-order autocorrelations in pattern recognition. *Information and Control*, Vol. 12, pp. 121{142, 1968.
- [Nar77] Eugene Narmour. *Beyond Schenkerism : the need for alternatives in music analysis*. University of Chicago Press, 1977.
- [NK89] Masako Nishijima and Yuji Kijima. Learning on sense of rhythm with a neural network - the Neuro Drummer. *Proceedings of the 1989 International Conference on Music Perception and Cognition*, pp. 78{80, October 1989.
- [NW92] Masako Nishijima and Kazuyuki Watanabe. Interactive music composer based on neural networks. *Proceedings of the 1992 International Computer Music Conference*, pp. 53{56, October 1992.
- [PSR93] Bruce Pennycook, Dale R. Stammen, and Debbie Reynolds. Toward a computer model of a jazz improviser. *Proceedings of the 1993 International Computer Music Conference*, pp. 228{231, September 1993.
- [RN95] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1995.
- [Row92] Robert Rowe. Machine listening and composing with Cypher. *Computer Music Journal*, Vol. 16, No. 1, pp. 43{63, Spring 1992.
- [Slo87] Jhon A. Sloboda. 演奏. 寺西立年, 大串健吾, 宮崎謙一 (編), *ダイアナ・ドイチュ 音楽の心理学 (下)*, 第 16 章, pp. 588{609. 西村書店, November 1987.
- [SP93] Dale R. Stammen and Bruce Pennycook. Real-time recognition of melodic fragments using the dynamic timewarp algorithm. *Proceedings of the 1993 International Computer Music Conference*, pp. 232{235, September 1993.
- [SP94] Dale R. Stammen and Bruce Pennycook. Real-time segmentation of music using an adaptation of Lerdahl and Jackendo®'s grouping principles.

- Proceedings of the 1994 International Conference of Music Perception and Cognition, pp. 269{270, July 1994.
- [ST] Daniel Sleator and David Temperley. The melisma music analyzer. Web: <http://www.link.cs.cmu.edu/music-analysis/>.
- [Tem01] David Temperley. The Cognition of Basic Musical Structures. MIT Press, November 2001.
- [Wey02] Tillman Weyde. Integrating segmentation and similarity in melodic analysis. Proceedings of the 2002 International Conference of Music Perception and Cognition, pp. 240{243, July 2002.
- [Wid94] Gerhard Widmer. Learning expression at multiple structural levels. Proceedings of the 1994 International Computer Music Conference, pp. 95{101, September 1994.
- [Wid95] Gerhard Widmer. Modeling the rational basis of musical expression. Computer Music Journal, Vol. 19, No. 2, pp. 76{96, Summer 1995.
- [バー 97] バーン・コーポレーション (編). BURRN! No. 175. (株) バーン・コーポレーション, September 1997.
- [プレ 98] プレイヤー・コーポレーション. Young Mates Music Player (月刊 YMM プレイヤー). (株) プレイヤー・コーポレーション, August 1998.
- [リッ 98] リットーミュージック (編). ギターマガジン (Guitar magazine), 第 19 卷 of 8. (株) リットーミュージック, August 1998.
- [井田 01] 井田健太郎, 平田圭二, 東条敏. GTTM に基づくグルーピング構造及び拍節構造の自動分析の試み. 情報処理学会研究報告, 2001-MUS-42, Vol. 2001, No. 103, pp. 49{54, October 2001.
- [加藤 89] 加藤俊一, 下垣弘行, 藤村是明. 画像対話型商標・意匠データベース TRADEMARK. 電子情報通信学会論文誌, Vol. J72-D-II, No. 4, pp. 535{544, April 1989.

- [加藤 92] 加藤俊一, 栗田多喜夫. 画像の内容検索電子美術館への応用. 情報処理, Vol. 33, No. 5, pp. 466{477, May 1992.
- [橋田 02] 橋田光代, 野池賢二, 平賀瑠美, 平田圭二, 片寄晴弘. FIT2002-Rencon Workshop -報告と課題, . 情報処理学会研究報告 2002-MUS-48, Vol. 2002, No. 123, pp. 35{39, December 2002.
- [近藤 93] 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, Vol. 1-373, No. 7Q-8, March 1993.
- [金森 95] 金森務, 片寄晴弘, 新美康永, 平井宏, 井口征士. ジャズセッションシステムのための音楽認識処理の一実現法. 情報処理学会論文誌, Vol. 36, No. 1, pp. 139{152, January 1995.
- [栗田 92] 栗田多喜夫, 加藤俊一, 福田郁美, 坂倉あゆみ. 印象語による絵画データベースの検索. 情報処理学会論文誌, Vol. 33, No. 11, pp. 1373{1383, November 1992.
- [栗田 93] 栗田多喜夫. 柔らかな情報処理のための統計的手法の応用. 電子技術総合研究所研究報告, No. 957, pp. 126{150, November 1993.
- [後藤 98a] 後藤真孝. 音楽音響信号を対象としたリアルタイムビートトラッキング. 人工知能学会研究会資料 AI チャレンジ研究会, Vol. SIG-Challenge-9801-2, pp. 7{14, November 1998.
- [後藤 98b] 後藤真孝. 拍節認識 (ビートトラッキング). bit 別冊 コンピュータと音楽の世界 | 基礎からフロンティアまで | , pp. 100{116. 共立出版, August 1998.
- [後藤 99] 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, No. 4, pp. 1910{1921, April 1999.

- [後藤 01] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一. RWC 研究用音楽データベース: ポピュラー音楽データベースと著作権切れ音楽データベース. 情報処理学会研究報告 2001-MUS-42, Vol. 2001, No. 103, pp. 35{42, October 2001.
- [後藤 02a] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一. RWC 研究用音楽データベース: クラシック音楽データベースとジャズ音楽データベース. 情報処理学会研究報告 2002-MUS-44, Vol. 2002, No. 14, pp. 25{32, February 2002.
- [後藤 02b] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一. RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース. 情報処理学会研究報告 2002-MUS-45, Vol. 2002, No. 40, pp. 19{26, May 2002.
- [今田 95] 今田達(編). 孤高のジプシーギター ジャンゴ・ラインハルト, 読んで聴く CD マガジン JAZZ, 第9巻. (株)同朋舎出版, January 1995.
- [小田 95] 小田安彦, 白川健一, 村上豊, 梶川嘉延, 野村康雄. 演奏者情報を加味したピアノの自動演奏システムの構築~音符属性に着目した個人的解釈ルールの生成~. 情報処理学会研究報告 95-MUS-10, Vol. 95, No. 46, pp. 7{12, October 1995.
- [上符 98] 上符祐一. コーパスを用いた音楽構造解析. 大阪大学 基礎工学部 修士論文, 1998.
- [青野 97] 青野裕司, 片寄晴弘, 井口征士. 重回帰分析を用いた演奏表現法の抽出. 情報処理学会論文誌, Vol. 38, No. 7, pp. 1473{1481, July 1997.
- [青野 99] 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, Vol. J82-DII, No. 11, pp. 1847{1856, November 1999.
- [石川 02] 石川修, 片寄晴弘, 井口征士. 重回帰分析のイタレーションによる演奏ルールの抽出と解析. 情報処理学会論文誌, Vol. 43, No. 2, pp. 268{276, February 2002.
- [浅野 90] 浅野哲夫. 計算幾何学. 朝倉書店, September 1990.

- [大津 81] 大津展之. パターン認識における特徴抽出に関する数理的研究. 電子技術総合研究所研究報告, No. 818, pp. 177{180, July 1981.
- [大槻 02] 大槻知史, 齋藤直樹, 中井満, 下平博, 嵯峨山茂樹. 隠れマルコフモデルによる音楽リズムの認識. 情報処理学会論文誌, Vol. 43, No. 2, pp. 245{254, February 2002.
- [東洋 02] 東洋武士, 平田圭二, 東条敏, 佐藤健. グルーピング規則適用を改良した GTTM の実装. 情報処理学会研究報告, 2002-MUS-47, Vol. 2002, No. 100, pp. 121{126, October 2002.
- [日高 96] 日高伊佐夫, 後藤真孝, 村岡洋一. すべてのプレイヤーが対等なジャズセッションシステム II, 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, February 1996.
- [白川 95] 白川健一, 小田安彦, 熊谷俊行, 梶川嘉延, 野村康雄. 演奏者情報を加味したピアノの自動演奏システムの構築 ~ ニューラルネットワークによる局所部分に於ける演奏特徴の抽出 ~. 情報処理学会研究報告 95-MUS-10, Vol. 95, No. 46, pp. 7{12, October 1995.
- [平賀 02] 平賀瑠美, 平田圭二, 片寄晴弘. 蓮根: めざせ世界一のピアニスト. 情報処理, Vol. 43, No. 2, pp. 136{141, February 2002.
- [平田 97] 平田圭二. IJICAI'97 ワークショップ「AI と音楽における課題 - 評価」の開催報告. 情報処理学会研究報告 97-MUS-22, Vol. 97, No. 100, pp. 19{24, October 1997.
- [平田 01a] 平田圭二, 後藤真孝. 音楽システムを考える. 情報処理学会研究報告 2001-MUS-40, Vol. 2001, No. 45, pp. 47{54, March 2001.
- [平田 01b] 平田圭二, 平賀瑠美. ハーヒーフン: 2 段階演奏表情付け法によるインクリメンタルな演奏生成システム. 情報処理学会研究報告 2001-MUS-39, Vol. 2001, No. 16, pp. 19{25, February 2001.

- [平田 02] 平田圭二, 青柳龍也. 音楽理論 GTTM に基づく多声音楽の表現手法と基本演算. 情報処理学会論文誌, Vol. 43, No. 2, pp. 277{286, February 2002.
- [片寄 90] 片寄晴弘, 井口征二. 知的採譜システム. 人工知能学会誌, Vol. 5, No. 1, pp. 59{66, January 1990.
- [片寄 02a] 片寄晴弘, 平賀瑠美, 平田圭二, 野池賢二. 蓮根 Performance Rendering Concours for Piano) について-System WG の活動を中心として-. 情報処理学会研究報告 2002-MUS-44, Vol. 2002, No. 14, pp. 19{24, February 2002.
- [片寄 02b] 片寄晴弘, 平賀瑠美, 平田圭二, 野池賢二, 橋田光代. ICAD-RENCON-報告と課題-. 情報処理学会研究報告 2002-MUS-47, Vol. 2002, No. 100, pp. 79{83, October 2002.
- [北 99] 北研二. 隠れマルコフモデル. 辻井潤一 (編), 確率的言語モデル, 言語と計算 4, 第 4 章, pp. 101{125. 東京大学出版会, November 1999.
- [林 88] 林知己夫. データ解析法の進歩. 放送大学教育振興会, March 1988.
- [鈴木 99] 鈴木一史, 加藤俊一, 築根秀男. 主観的類似度に適応した 3次元多面体の検索. 電子情報通信学会論文誌, Vol. J82-D-I, No. 1, pp. 184{192, January 1999.
- [鈴木 00] 鈴木泰山, 徳永建伸, 田中穂積. 事例に基づく演奏表情の生成. 情報処理学会論文誌, Vol. 41, No. 4, pp. 1134{1145, April 2000.
- [和気 94] 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, July 1994.

著者論文リスト

論文誌

- [1] 浜中雅俊, 後藤真孝, 麻生英樹, 大津展之: "発音時刻の楽譜上の位置を確率モデルにより推定するクオンタイズ手法", 情報処理学会論文誌, Vol.43, No.2, pp.234-244, February 2002. (4.2節の内容に相当)
- [2] 浜中雅俊, 後藤真孝, 麻生英樹, 大津展之: "演奏者の振る舞いを統計的に学習するジャムセッションシステム", 情報処理学会論文誌. (投稿中, 3章の内容に相当)
- [3] 上江洲弘明, 津田 栄, 箭内美智子, 浜中雅俊, 稲井田次郎: "ファジィノードファジィグラフ解析とその応用", 早稲田大学数学教育学会論文誌, Vol.19, No.1, pp.43-47, November 2001.
- [4] 浜中雅俊, 稲井田次郎: "ファジィグラフを応用した曲順の設定法", 早稲田大学数学教育学会論文誌, Vol.14, No.1, pp.34-39, November 1996. (ショートノート)

国際会議等

- [5] Masatoshi Hamanaka, Masataka Goto, Nobuyuki Otsu: "Learning-Based Jam Session System for a Guitar Trio, Proceedings of the 2001 International Computer Music Conference," pp.467-470, Havana, Cuba, September 2001. (査読つき 採択率 32%, 3章の内容に相当)
- [6] Masatoshi Hamanaka, Masataka Goto, Hideki Asoh, Nobuyuki Otsu: "A Learning-Based Quantization: Estimation of Onset Times in a Musical Score," World Multiconference on Systemics, Cybernetics and Informatics 2001, Vol.X,

pp.374-379, Orland, USA, July 2001. (査読つき 採択率 63 % , 4.2 節の内容の一部に相当)

- [7] Masatoshi Hamanaka, Jiro Inaida, Akihiko Yanai, Michiko Yanai, Ei Tsuda, Hajime Yamashita: "Editing a Music for Smooth Playing Order Applying Fuzzy Graph", Proceedings of the Eighth International Fuzzy System Association World Congress, Vol. , pp.1024-1028, Taipei, Taiwan ROC, August 1999. (査読つき)
- [8] Masatoshi Hamanaka and Keiji Hirata, Applying Voronoi Diagram in the Automatic Grouping of Polyphony, Information Technology Letters (FIT2002 情報技術レターズ), Vol.1, No.1, pp.101-102, Tokyo, September 2002. (査読つき 採択率 34 % , 4.3.3 の内容に相当)
- [9] Masatoshi Hamanaka, Jiro Inaida, Akihiko Yanai, Michiko Yanai, Ei Tsuda, Hajime Yamashita: "Natural Edition of Music Sequence Applying Fuzzy Graph", Proceedings of the Fifteenth Fuzzy System Symposium, Vol. , pp.455-458, Osaka, Japan, June 1999.

研究会

- [10] 浜中雅俊, 後藤真孝, 大津展之: "学習するジャムセッションシステム:演奏者固有のフレーズの獲得", 情報処理学会 音楽情報科学研究会 研究報告, 2002-MUS-47-13, Vol.2002, No.100, pp.71-78, 金沢, October 2002. (4 章の内容の一部に相当)
- [11] 浜中雅俊, 後藤真孝, 麻生英樹, 大津展之: "学習に基づくクオンタイズ:発音時刻の楽譜上の位置の推定", 情報処理学会 音楽情報科学研究会 研究報告, 2001-MUS-40-4, Vol.2001, No.45, pp.21-28, 東京, May 2001. (4.2 節の内容の一部に相当)
- [12] 浜中雅俊, 後藤真孝, 大津展之: "学習するジャムセッションシステム:演奏者の振る舞いのモデルの獲得", 情報処理学会 音楽情報科学研究会 研究報告, 2000-MUS-34-5, Vol.2000, No.19, pp.27-34, 東京, February 2000. (3 章の内容に相当)

- [13] 浜中雅俊, 築根秀男: "ドラムスの生演奏に MIDI 音源の再生音をリアルタイムでシンクロさせるシステムの開発", 情報処理学会 音楽情報科学研究会 研究報告, 98-MUS-26-8, Vol.98, No.74, pp.53-60, 北海道, May 1998.

全国大会・シンポジウム

- [14] 浜中雅俊, 稲井田次郎, 箭内顯彦, 箭内美智子, 奥田良治, 山下元: "ファジィ理論を応用したギター運指経路の選択法", 数学教育学会 春季年会研究紀要, pp.161-163, 東京, March 1998.
- [15~17] 浜中雅俊, 稲井田次郎, 箭内顯彦, 箭内美智子, 奥田良治, 山下元: "ファジィグラフを応用した曲順の編集法 I~III", ファジィシステムシンポジウム, 1996~1998.
- [18~21] 浜中雅俊, 稲井田次郎, 箭内顯彦, 箭内美智子, 奥田良治, 山下元: "ファジィ理論を応用した曲順の編集法 ~IV", 数学教育学会紀要, 1996~1998.

その他の口頭発表

- [22] 浜中雅俊, 後藤真孝, 麻生英樹, 大津展之: "確率モデルに基づくクォンタイズ: ゆらぎを含む演奏から量子化された楽譜上の発音時刻系列の推定", 「感性・ファジィ」合同シンポジウム, pp.44-51, 東京, March 2002.
- [23] 浜中雅俊, 後藤真孝: "なりかわりセッションシステム - 人間の演奏者の振る舞いを学習した仮想演奏者との共演 - :", 筑波大学 感性評価構造モデル構築特別プロジェクト 研究報告, Vol.3, pp.213-218, 茨城, September 2000.
- [24] 浜中雅俊: "ICMC2001 報告", 情報処理学会 音楽情報科学研究会 研究報告, 2001-MUS-43-7, Vol.2001, No.125, pp.39-45, 神奈川, December 2001.
- [25] 浜中雅俊, 稲井田次郎, 箭内顯彦, 箭内美智子, 奥田良治, 山下元: "ファジィ理論を応用した音曲の系列分析", 日本ファジィ学会ノンエンジニアリング・ファジィ研究会 ワークショップ, pp.86-89, 愛知, March 1997.

特記事項

- 2 論文 12 により，情報処理学会 平成 13 年度山下記念研究賞 受賞．
- 2 論文 6 により，SCI (World Multi-conference on Systemics, Cybernetics and Informatics) in Art ベストペーパー賞 受賞．

付録A 主成分分析による印象語選択

3.3.1では、印象ベクトルの次元数を少なくするために、印象ベクトルを主成分分析し、重要な印象語を選択した。付録Aでは、その具体的な手法について説明する。

心理実験から得られたデータは、表A.1のようにまとめられる。 $x_1; x_2; \dots; x_n$ は、 $p(= 10)$ 個の印象語それぞれに対する評定結果である。 $n(= 1440)$ 個は標本数である。

表 A.1: 心理実験から得られたデータ

標本番号	変量									
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	$x_{1\ 1}$	$x_{1\ 2}$	$x_{1\ 3}$	$x_{1\ 4}$	$x_{1\ 5}$	$x_{1\ 6}$	$x_{1\ 7}$	$x_{1\ 8}$	$x_{1\ 9}$	$x_{1\ 10}$
2	$x_{2\ 1}$	$x_{2\ 2}$	$x_{2\ 3}$	$x_{2\ 4}$	$x_{2\ 5}$	$x_{2\ 6}$	$x_{2\ 7}$	$x_{2\ 8}$	$x_{2\ 9}$	$x_{2\ 10}$
:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:
n	$x_{n\ 1}$	$x_{n\ 2}$	$x_{n\ 3}$	$x_{n\ 4}$	$x_{n\ 5}$	$x_{n\ 6}$	$x_{n\ 7}$	$x_{n\ 8}$	$x_{n\ 9}$	$x_{n\ 10}$
平均	\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4	\bar{x}_5	\bar{x}_6	\bar{x}_7	\bar{x}_8	\bar{x}_9	\bar{x}_{10}

このとき、第1主成分は、 p 個の変量の1次結合 $z = l_1x_1 + l_2x_2 + \dots + l_px_p$ の $l_1; l_2; \dots; l_p$ を変化させ、 $l_i^2 = 1$ の条件の下で、 z の分散が最大となる z である。

そして、第 i 主成分は、 $z_1; z_2; \dots; z_{i-1}$ と無相関で、分散が最大になる z_i である。具体的方法を以下に示す。

各変量に対して、偏差積和を計算すると、偏差積和行列が得られる。

$$A = \begin{pmatrix} a_{1\ 1} & a_{1\ 2} & \dots & a_{1\ 10} \\ a_{2\ 1} & a_{2\ 2} & \dots & a_{2\ 10} \\ \vdots & \vdots & \ddots & \vdots \\ a_{10\ 1} & a_{10\ 2} & \dots & a_{10\ 10} \end{pmatrix} \quad (\text{A.1})$$

ただし，

$$a_{ij} = \sum_{s=1}^X (x_{s,i} - \bar{x}_i)(x_{s,j} - \bar{x}_j) \tag{A.2}$$

このとき，分散共分散行列 S は，式 A.3 で得られる．

$$S = \begin{pmatrix} \frac{3}{4} s_{11} & \frac{3}{4} s_{12} & \dots & \frac{3}{4} s_{10} \\ \frac{3}{4} s_{21} & \frac{3}{4} s_{22} & \dots & \frac{3}{4} s_{20} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{3}{4} s_{101} & \frac{3}{4} s_{102} & \dots & \frac{3}{4} s_{1010} \end{pmatrix} \tag{A.3}$$

ただし，

$$\frac{3}{4} s_{ij} = a_{ij} / (n - 1) \tag{A.4}$$

したがって，固有方程式 $|S - \lambda I| = 0$ は，式 A.5 で与えられる．

$$\begin{vmatrix} \frac{3}{4} s_{11} - \lambda & \frac{3}{4} s_{12} & \dots & \frac{3}{4} s_{10} \\ \frac{3}{4} s_{21} & \frac{3}{4} s_{22} - \lambda & \dots & \frac{3}{4} s_{20} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{3}{4} s_{101} & \frac{3}{4} s_{102} & \dots & \frac{3}{4} s_{1010} - \lambda \end{vmatrix} = 0 \tag{A.5}$$

この固有方程式を解き，固有値 $(\lambda_1, \lambda_2, \dots, \lambda_9)$ が得られる．各主成分 Z_{\circ} の寄与率 C_{\circ} は，式 A.6 で求められる．

$$C_{\circ} = \frac{\lambda_{\circ}}{\sum_{i=1}^X \frac{3}{4} s_{ii}} \tag{A.6}$$

その結果，累積寄与率が 84 % までの主成分が 3 つ得られた．第 3 主成分までの固有値と累積寄与率を表 A.2 に示す．

固有値 λ_{\circ} を $|S - \lambda_{\circ} I|$ に代入し， $|l^T l| = 1$ となるような固有ベクトル $l_{\circ i}$ を求める．このとき，主成分と x_i との相関は，因子負荷量 $r_{a,i}$ で表される（式 A.7）．

$$r_{a,i} = \frac{\lambda_{\circ} l_{\circ i}}{\frac{3}{4} s_{ii}} \tag{A.7}$$

因子負荷量を調べたところ，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ることがわかった（表 A.3）．

表 A.2: 固有値と累積寄与率

主成分	第 1 主成分 z_1	第 2 主成分 z_2	第 3 主成分 z_3
固有値	1.343	0.931	0.727
累積寄与率	0.374	0.633	0.835

表 A.3: 存在感，躍動感，重厚感の因子負荷量

x_i	因子負荷量		
	第 1 主成分 z_1	第 2 主成分 z_2	第 3 主成分 z_3
x_8 (存在感)	0.826	0.055	-0.113
x_{10} (存在感)	0.053	0.835	-0.034
x_6 (重厚感)	-0.111	-0.035	0.827

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．そこで，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 次元で表すことにした．

付録B 特徴量抽出のための予備実験

3.3.1では、音量フィルタ、音域フィルタ、Noteフィルタという3種類のフィルタを用意し、様々な物理特徴量を分離、抽出することによって、663552個の物理特徴量を用意した。付録Bでは、3種類のフィルタを設計するために行った、どのような物理特徴量が、演奏から受ける印象に影響を与えているかを求める予備実験について説明する。

被験者は4コーラスのセッションの演奏を聴き、その演奏がソロであるか、伴奏であるかを判断する。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演奏記録の、ある1人分のパートの演奏である。判断は1小節の1/12の時間分解能で行う。

被験者がソロだと判断した部分と、伴奏だと判断した部分のMIDIデータのベロシティー（音量）、音高、コードの主音からの度数、ピッチベンドを比較した。

ベロシティーは、0から127の整数で表される。それを7等分し、それぞれの範囲のソロの音の回数と伴奏の音の回数を比較した（図B.1）。その結果、伴奏部分のベロシティーは90以下の場合が多く、ソロ部分のベロシティーは91以上の場合が多いことがわかった。そこで、音量フィルタには、ベロシティーが90以下の時に1

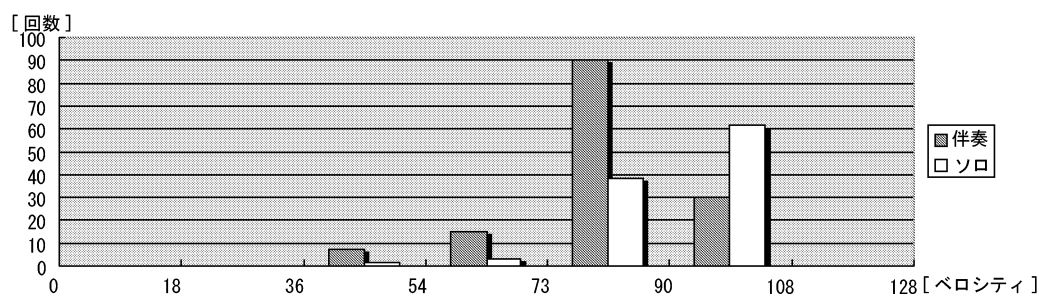


図 B.1: ソロと伴奏でのベロシティーの比較

を出力する関数，ベロシティーが 91 以上の時に 1 を出力する関数，常に 1 を出力する関数の 3 通りを用意した。

音高は，1 から 127 の整数で表される．それを 7 等分し，それぞれの範囲のソロの音の回数と伴奏の音の回数を比較した（図 B.2）．その結果，伴奏部分の音高は，54 以下の場合が多く，ソロ部分の音高は 55 以上の場合が多いことがわかった．そこで，音域フィルタには，音高が 54 以下の場合に 1 を出力する関数，音高が 55 以上の場合に 1 を出力する関数，常に 1 を出力する関数の 3 通りを用意した。

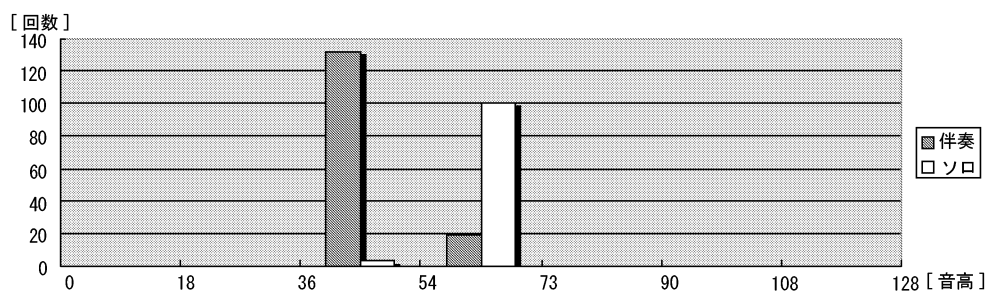


図 B.2: ソロと伴奏での音高の比較

コードの主音からの度数は，1 から 13 までの整数で表される．それぞれの度数のソロの音の回数と伴奏の音の回数を比較した（図 B.3）．その結果，11 度，13 度等の音は，伴奏の時に多用され，2 度，10 度等の音は，ソロの時に多用されることがわかった．Note フィルタは，コードの 2 度または 10 度の音が出ていたら 1 を出力する関数など，全部で 512 通りを用意した。

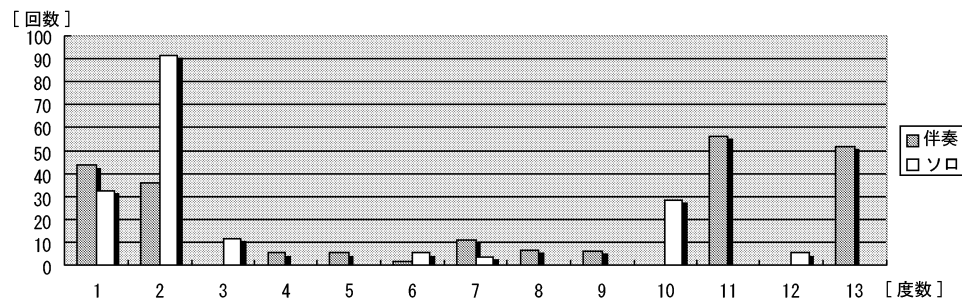


図 B.3: ソロと伴奏でのコードの主音からの度数の比較

ピッチベンドは、-8191 から +8191 までの整数で表される。ここでは、それを 45 等分し、それぞれの範囲でのソロのピッチベンド回数と、伴奏のピッチベンド回数を比較した (図 B.4)。ピッチベンド回数は、ソロと伴奏でほとんど違いが認められなかった。

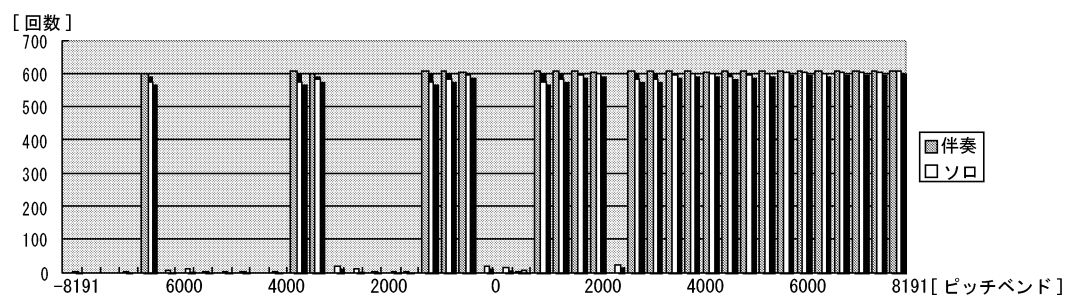


図 B.4: ソロと伴奏でのピッチベンドの比較

以上の予備実験の結果により、演奏の物理特徴の違いから、ある程度ソロと伴奏の判断が可能であるとの知見が得られた。本研究では、予備実験により得られた知見を利用して物理特徴量を抽出しているが、システム自体ではソロと伴奏の明示的な判断をしていない。本研究では、ソロと伴奏の判断は、物理特徴量から直接得られるものではなく、演奏から受ける印象をもとに判断していると考えている。

付 録 C 正 準 相 関 分 析

3.3.1 では、印象ベクトルと相関の高い物理特徴量を正準相関分析で求めた。付録 C では、正準相関分析の具体的な手法について説明する。

いま、 $s(= 3)$ 個の印象語の評定データを第 1 組の変数とし、 $t(= 100)$ 個の物理特徴量から得られたデータを第 2 組の変数群とする (表 C.1)。 $n(= 1440)$ 個は、標本数である。

表 C.1: 印象ベクトルと物理特徴ベクトルのデータ

標本番号	第 1 組の変数 (印象語の評定値)			第 2 組の変数 (物理特徴量)					
	x_1	x_2	x_3	x_4	x_5	x_6	...	x_{102}	x_{103}
1	$x_{1\ 1}$	$x_{1\ 2}$	$x_{1\ 3}$	$x_{1\ 4}$	$x_{1\ 5}$	$x_{1\ 6}$...	$x_{1\ 102}$	$x_{1\ 103}$
2	$x_{2\ 1}$	$x_{2\ 2}$	$x_{2\ 3}$	$x_{2\ 4}$	$x_{2\ 5}$	$x_{2\ 6}$...	$x_{2\ 102}$	$x_{2\ 103}$
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
n	$x_{n\ 1}$	$x_{n\ 2}$	$x_{n\ 3}$	$x_{n\ 4}$	$x_{n\ 5}$	$x_{n\ 6}$...	$x_{n\ 102}$	$x_{n\ 103}$
平均	\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4	\bar{x}_5	\bar{x}_6	...	\bar{x}_{102}	\bar{x}_{103}

ここでは、記述の簡単のため、印象ベクトルと物理特徴ベクトルのデータは、

$$\begin{matrix}
 \mathbf{X} \\
 x_{i\ j} = 0 \quad \text{ただし } (i = 1; 2; \dots; s + t)
 \end{matrix}
 \tag{C.1}$$

$i=1$

となるように正規化してあるものとする。このとき、2 組の変数群の線形結合

$$\begin{aligned}
 u_k &= l_{k\ 1}x_1 + l_{k\ 2}x_2 + \dots + l_{k\ s}x_s = l_k^0 x_1 \quad (k = 1; 2; \dots; s) \tag{C.2} \\
 v_k &= m_{k\ 1}x_{s+1} + m_{k\ 2}x_{s+2} + \dots + m_{k\ s}x_{s+t} = m_k^0 x_2 \quad (k = 1; 2; \dots; t)
 \end{aligned}$$

の相関が最大となるような、 u_k, v_k は、正準相関分析により求まる。以下、具体的な計算方法について述べる。

まず、表 C.1 から、相関行列 R を求める。

$$R = \begin{matrix} & \begin{matrix} X_1 & X_2 & X_3 & X_4 & \dots & X_{103} \end{matrix} \\ \begin{matrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{matrix} & = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & \dots & r_{1103} \\ r_{21} & r_{22} & r_{23} & r_{24} & \dots & r_{2103} \\ r_{31} & r_{32} & r_{33} & r_{34} & \dots & r_{3103} \\ \hline r_{41} & r_{42} & r_{43} & r_{44} & \dots & r_{4103} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ r_{1031} & r_{1032} & r_{1033} & r_{1034} & \dots & r_{103103} \end{bmatrix} \end{matrix} \quad (C.3)$$

このとき、

$$A = R_{12} R_{22}^{-1} R_{21} \quad (C.4)$$

$$B = R_{11}$$

とすると、正準変数、正準相関は、次の固有方程式を解くことにより求まる。

$$|A - \lambda B| = 0 \quad (C.5)$$

固有方程式の解を $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_s$ とすると、正準相関は、

$$r(k) = \sqrt{\lambda_k} \quad (k = 1; 2; \dots; s) \quad (C.6)$$

このとき、第1組の正準変数 $L (= [l_1; l_2; \dots; l_s])$ は、固有ベクトルを、

$$L^T R_{11} L = I \quad (C.7)$$

を満たすように基準化することにより求められる。一方、第2組の正準変数 $M (= [m_1; m_2; \dots; m_s])$ は、次式より求まる。

$$m_k = \frac{1}{r(k)} R_{22}^{-1} R_{21} l_k \quad (C.8)$$

このとき、正準変数 u_k と x_{s+1} との相関 $r(u_k; x_{s+1})$ は、次式より求まる。

$$\begin{matrix} r(u_k; x_{s+1}) \\ r(u_k; x_{s+2}) \\ \vdots \\ r(u_k; x_{s+t}) \end{matrix} = R_{22} m_k \quad (k = 1; 2; \dots; s) \quad (C.9)$$

以上のような計算により，100 個の物理特徴量のうち，正準変数と一番相関の低い物理特徴量を求め，それを削除し，新たな物理特徴量を 1 つ加えて再度正準相関分析を行うという操作を繰り返すことによって，重要な特徴量を取捨選択できる．

新たに追加する物理特徴量は，残っている 99 個の物理特徴量のいずれに対しても相関の低いものとした．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（式 C.3）の状態が悪くなり，式 C.4 で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．

付録D グラムシュミットの直交化法

3.3.2では、グラムシュミットの直交化法を用いて、RBFネットワークの出力に対して寄与の大きいユニットを選択した。付録Dでは、グラムシュミット法を用いて、重要な直交ベクトルを算出する計算手順について示す。

グラムシュミット法は、互いに独立なベクトル $P(x) = [p_1; p_2; \dots; p_M]; p_i \in \mathbb{R}^n$ が与えられたときに、そのベクトルが張る線形空間の直行基底を求める算法である。

以下に、グラムシュミット法の一般的な計算法を示す。

手順1 $w_1 = p_1$

手順2 $k = 2; 3; \dots; M$ について、以下の操作を繰り返す。

$$v_{ik} = \frac{w_i^T p_k}{w_i^T w_i} \quad (i = 1; \dots; k-1) \quad (D.1)$$

$$w_k = p_k - \sum_{i=1}^{k-1} v_{ik} w_i \quad (D.2)$$

以上のようにして、 $w_1; w_2; \dots; w_{k-1}$ に直交する w_k を、 $p_1; p_2; \dots; p_{k-1}$ から構成することができる。したがって、 p_i を重要なものから選択することにより、入力から出力への寄与の大きなユニットを使って、RBFネットワークを構成することができる。

しかし、グラムシュミット法自体には、重要な成分ベクトルを選ぶという機能はないため、それを実現するためには、以下のようなアルゴリズムを構築する必要がある。 $P(x), y_j$ は3.3.2の式3.3での定義と同じである。

手順1 ($k = 1$) $i = 1; 2; \dots; M$ について以下を計算する。

$$w_1^{(1)} = p_i \quad (D.3)$$

$$g_1^{(i)} = \frac{w_1^{(i)T} y}{w_1^{(i)T} w_1^{(i)}} \quad (D.4)$$

$$r_1^{(i)} = \frac{\mathbf{i} \mathbf{g}_1^{(i)} \mathbf{C}_2 \mathbf{i} \mathbf{w}_1^{(i)} \mathbf{C}_T \mathbf{w}_1^{(i)}}{y^T y} \quad (\text{D.5})$$

$$i_1 = \operatorname{argmax}_{1 \leq i \leq M} r_1^{(i)} \quad (\text{D.6})$$

$$\mathbf{u}_1 = \mathbf{w}_1^{(i_1)} = \mathbf{p}_{i_1} \quad (\text{D.7})$$

$$R(1) = r_{i_1} \quad (\text{D.8})$$

手順 k ($k = 2; 3; \dots; M$) それぞれの k に対して $i = 1; 2; \dots; M$ について, 以下を計算し (ただしすでに選ばれた $i = i_1; i_2; \dots; i_{k-1}$ については計算不要), $R(k)$ があらかじめ, 決めた値を超えれば終了, そうでないときは $k := k + 1$ として繰り返す.

$$v_{jk}^{(i)} = \frac{\mathbf{u}_j^T \mathbf{p}_i}{\mathbf{u}_j^T \mathbf{u}_j} \quad (\text{D.9})$$

$$\mathbf{w}_k^{(1)} = \mathbf{p}_i - \sum_{j=1}^{k-1} v_{jk}^{(i)} \mathbf{u}_j \quad (\text{D.10})$$

$$g_k^{(i)} = \frac{\mathbf{i} \mathbf{w}_k^{(i)} \mathbf{C}_T \mathbf{y}}{\mathbf{i} \mathbf{w}_k^{(i)} \mathbf{C}_T \mathbf{w}_k^{(i)}} \quad (\text{D.11})$$

$$r_k^{(i)} = \frac{\mathbf{i} \mathbf{g}_k^{(i)} \mathbf{C}_2 \mathbf{i} \mathbf{w}_k^{(i)} \mathbf{C}_T \mathbf{w}_k^{(i)}}{y^T y} \quad (\text{D.12})$$

$$i_k = \operatorname{argmax}_{1 \leq i \leq M; i \notin \{i_1, i_2, \dots, i_{k-1}\}} r_k^{(i)} \quad (\text{D.13})$$

$$\mathbf{u}_k = \mathbf{w}_k^{(i_k)} = \mathbf{p}_{i_k} - \sum_{j=1}^{k-1} v_{jk}^{i_k} \mathbf{u}_j \quad (\text{D.14})$$

$$R(k) = R(k-1) + r_{i_k} \quad (\text{D.15})$$

以上の手順 $1; 2; \dots; M$ により, 入力ベクトルから出力ベクトルへの寄与の大きい M 個のユニットが選択でき, RBF ネットワークを決定することができる.

付録E スタインのパラドックス (経験的ベイズ法)

4.2.3では、隠れマルコフモデルの出力分布の形状 $\hat{b}_i(l)$ を、平均的な分布 $\hat{b}_i(l)$ と最尤推定により得た分布 $b_i(l)$ との補間により推定した $\hat{b}_i(l)$ が、 $b_i(l)$ よりも良い推定値となるという考え方は、スタインのパラドックスに基づいている。付録Eでは、スタインのパラドックスについて具体的に説明する。

スタインの定理とは、3種類以上の独立な確率変数の期待値を推定する場合、その確率変数の観測値（繰り返し測定なら平均値）を推定値とするよりもすぐれた推定量が存在するという定理である。これは、観測された平均を用いる方法より一様により推定規則はないという古典的な統計理論の基本的な原理と矛盾していることから、スタインのパラドックスと呼ばれている。

ここでは、野球選手の打率の例を使ってスタインのパラドックス（経験的ベイズ法）について説明する。

いまシーズンの初めに、複数の選手の打率 y （安打数/打数）がわかっているとき、将来の打率を予測したいとする。このとき、各選手についてのジェームズ・スタイン推定量は次式で表される。

$$z = \bar{y} + c(y_i - \bar{y}) \quad (\text{E.1})$$

ここで、 \bar{y} は、すべての選手の打率の総平均である。したがって、 $(y_i - \bar{y})$ は、個々の選手の打率の平均が、総平均からどの程度離れているかを示す。 c は、縮減因子と呼ばれ、次の方程式で定義される。

$$c = 1 - \frac{(k-3)\frac{\sigma^2}{k}}{(y_i - \bar{y})^2} \quad (\text{E.2})$$

k は選手の数、 $\frac{\sigma^2}{k}$ は標準偏差の2乗である。

図E.1は、10人の選手について、ジェームズ・スタイン推定量を計算した結果である。この図からも明らかなように、ジェームズ・スタイン推定量は、個々の打率

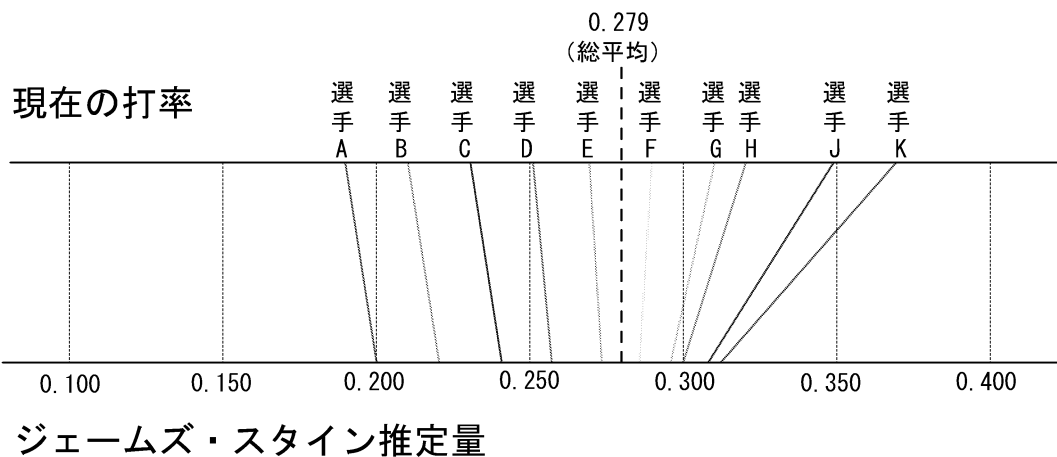


図 E.1: 10人の野球選手に対するジェームズ・スタイン推定量

を総平均へ向かって、縮減した値である。この場合、総平均は0.279であり、それぞれの平均は総平均への距離の15%から65%を縮めていることになる¹。スタインの方法の考え方の根拠は、理論平均というものは、データの各平均によって観察されるよりも、ずっと互いに似通ったものであると考えるところにある。これは、ベイズの定理の事前分布の考え方に近い。

今、全選手の打率の平均 m と標準偏差 C を求めると、それぞれ0.279, 0.057となる。これを事前分布とすると、ベイズ推定量 Z は、次式で表される。

$$Z = m + C(y_j - m) \quad (\text{E.3})$$

ジェームズ・スタイン推定量の式(式E.1)と、ベイズ推定量の式(式E.3)は、確率変数の数(打率を推定する野球選手の数)が大きくなる極限で両者は一致する。ジェームズ・スタインの方法がベイズの方法よりも有利な点は、事前分布の知識がなくても(すなわち分布が正規分布であるかわからない場合でも)行える手法であることである。

¹式E.2からわかるように、縮減因子は各選手の標準偏差によって異なる。標準偏差は、打率が5割に近づくほど高くなる。したがって、打率の低い選手ほど縮減因子が小さくなり、その結果縮減が小さい。

付録F MDA-OR(最小次元解析)

4.4.1 では、多次元尺度法の1種であるMDA-OR (Minimum Dimension Analysis of Ordered Class Belonging) を用いて、100個のフレーズを、一対比較実験の結果求めた主観的類似度に基づき、3次元空間上に配置した。付録Fでは、MDA-ORを用いた空間配置の具体的な手法について示す。

MDA-ORでは、まず空間の次元を1次元として配置し、その値を固定し、次元数が不足している場合にはその残差を計算し、1次元目のときと同じ手法で2次元目の解を求める。そして、2次元目の解でもまた適合度が十分でなければ、2次元までの値を固定し残差を求め、また同じ手法で次の次元の解を求める。このような次元の増やし方を階層的次元作成法という。その特長は、適合度をみながら次元を上げるかどうか検討でき、しかも前の結果が保持され構造変化をきたさずに次元が増えて精度が向上する点である。

まず、1次元の場合を考える。2つのフレーズ j と k の類似度を $d_{jk} = x_j \cdot x_k$ ¹ とする。このとき、 j と k との関係 R_{jk} を、順序のついた G 個のクラスに分類する。そして $\pm_{jk}(g)$ を以下のように定義する。

$$\pm_{jk}(g) = 1; \quad R_{jk} \text{が } g \text{ 番目}^2 \text{のクラスに属しているとき} \quad (\text{F.1})$$

$$= 0; \quad \text{その他} \quad (\text{F.2})$$

$$\text{ただし } j, k = 1, 2, \dots, N; g = 1, 2, \dots, G$$

このとき、次式が成立する。T は、組み合わせの数で、欠測組み合わせがないときには、 $N(N-1)$ に等しい。

$$\sum_{g=1}^G \sum_{j \neq k} \pm_{jk}(g) = 1 \quad (\text{F.3})$$

¹ d_{jk} はユークリッドでも非ユークリッドでも構わない。

$$\sum_j \sum_{k(j \neq k)} \sum_g \pm_{jk}(g) = T \tag{F.4}$$

全分散 \sum_{jk}^2 は次式で表される.

$$\sum_{jk}^2 = \frac{1}{T} \sum_j \sum_{k(j \neq k)} d_{jk}^4 \sum_i Q^2 \tag{F.5}$$

ただし,

$$Q = \frac{1}{T} \sum_j \sum_{k(j \neq k)} d_{jk}^2 \tag{F.6}$$

内分散 $\sum_{jk}^2_w$ と外分散 $\sum_{jk}^2_b$ は次のようになる.

$$\sum_{jk}^2_w = \sum_g \frac{f_g}{T} \frac{1}{f_g} \sum_j \sum_{k(j \neq k)} d_{jk}^4 \pm_{jk}(g) \sum_i Q_g^2 \tag{F.7}$$

$$\sum_{jk}^2_b = \sum_g \frac{1}{f_g} Q_g^2 \sum_i Q^2 \tag{F.8}$$

ただし,

$$Q_g = \frac{1}{f_g} \sum_j \sum_{k(j \neq k)} d_{jk}^2 \pm_{jk}(g)$$

$$f_g = \sum_j \sum_{k(j \neq k)} \pm_{jk}(g)$$

このとき, 相関比 r^2 は, R_{jk} に対する d_{jk} の適合のよさを表す.

$$r^2 = \frac{\sum_{jk}^2_b}{\sum_{jk}^2} \tag{F.9}$$

上式において r^2 は x の関数である. そこで, 線形計画法の繰り返し計算により, x を求めることができる.

次に2次元の場合では, 2つのフレーズ j と k の類似度を $d_{jk} = (x_j - x_k)^2 + (y_j - y_k)^2$ とする. それから, 1次元のときと同じように分散, 内分散, 外分散を計算し, 相関比 r^2 を x と y の関数で表す. そして, 線形計画法により y を求める.

以上のような手順を次元を上げながら繰り返し, r^2 がある値を超えたらそこで終了とする. r^2 が最終的な適合度である.