



修 士 論 文

ドラム演奏のグルーヴ感の解析

Analysis of Groove Feelings of Drums Plays

指導教員 近山 隆 教授

東京大学大学院 新領域創成科学研究科
基盤情報学専攻

学生証番号 47 - 56340

氏 名 渡辺 哲朗

提 出 日 平成 19 年 1 月 31 日

概要

今日のコンピュータ音楽においては、本来の人間の演奏に内在するずれやゆらぎといった「グルーブ感」を表現することに困難がある。本研究では、プロのドラム奏者が実際に演奏したドラム演奏を収録した音源から打点時刻のずれと打点音量を抽出し、これらを Support Vector Machine (SVM) を用いて分類することにより、演奏に内在するグルーブ感の違いに対する定量的な解析を行った。

まず、打点時刻ずれ・打点音量それぞれの平均・標準偏差を各次元の値とする4次元の入力ベクトル群による特徴空間において、線形SVMおよびカーネルトリックを用いた非線形SVMによって分類を行い、「ルース」と「タイト」という2種類のグルーブ感を分類するための分離超平面・分離超曲面を生成した。その結果、線形分類・非線形分類ともに各楽器・各打拍の各平均正解率の大部分が80%台後半～100%という高い水準でそれぞれ分類を行うことができた。線形SVMによる解析では、特に「打点時刻ずれの標準偏差」がグルーブ感を分類する最も重要な要因となっている特徴であることが判明し、その次に重要であると判明した特徴は「打点音量の平均」であった。また、線形SVMによる分類と非線形SVMによる分類では正解率の差はほとんどなく、4次元特徴空間であれば線形分類であっても十分な分類が行えることが示された。

さらに、線形SVMによる解析で重要と判別された上記2種の特徴について、2次元特徴空間において非線形SVMによって分類を行い、各楽器・各打拍におけるグルーブ感の分離曲線を生成し、グルーブ感の分布を図示することができた。その結果、分類の正解率は4次元特徴空間のときと比べて低いものとなったが、ルースなグルーブ感の演奏に比べてタイトなグルーブ感の演奏の方が全般的に演奏内の打点のばらつきが少なく、また複数回の演奏を行った際にも安定した演奏が行われていることが示された。

キーワード：音楽情報処理，感性情報処理，機械学習

目次

第 1 章	序論	1
1.1	背景	1
1.1.1	コンピュータ音楽	1
1.1.2	コンピュータによるグルーブ感の再現	3
1.1.3	ドラムによるグルーブ感	4
1.2	本研究の目的	4
1.3	本論文の構成	5
第 2 章	関連研究	6
2.1	ドラムに関する基礎知識	6
2.1.1	ドラムの構成	6
2.1.2	演奏パターン	7
2.1.3	ドラム演奏の実際	9
2.2	人間のドラム演奏に関する研究	11
2.2.1	ドラム演奏に対する定量的解析	11
2.2.2	ドラム演奏のグルーブ感に関する研究	12
2.3	Support Vector Machine	16
2.3.1	線形 SVM による分類問題の解法	16
2.3.2	ソフトマージンアルゴリズム	19
2.3.3	カーネルトリック — 非線形分類への拡張	21
2.3.4	SVM の利点	23
第 3 章	提案手法	24
3.1	本研究のアプローチ	24

3.2	提案手法の概要	25
3.3	解析対象の音源データ	28
3.4	SVM 使用の背景	32
3.5	打点検出方法	32
3.5.1	ヒルベルト変換による包絡曲線の抽出	32
3.5.2	包絡曲線の平滑化処理	35
3.5.3	極大値探索による打点の検出	37
3.6	仮想メトロノーム	37
3.7	クロスバリデーションによる SVM パラメータチューニング	39
3.8	線形分離超平面の係数による重要な特徴の判別	41
第 4 章	SVM によるグルーブ感の解析	43
4.1	実験環境	43
4.2	線形 SVM による 4 次元特徴空間での分類と重要な特徴の判別	44
4.2.1	実験手順	44
4.2.2	入力データの次元について	45
4.2.3	実験結果	45
4.2.4	考察 — 重要な特徴の判別	55
4.3	非線形 SVM による 4 次元特徴空間での分類と分類性能比較	57
4.3.1	実験手順	57
4.3.2	入力データの次元について	58
4.3.3	カーネル関数の選定	58
4.3.4	実験結果	59
4.3.5	考察 — 線形分類と非線形分類の性能比較	61
4.4	非線形 SVM による 2 次元特徴空間での分類	63
4.4.1	実験手順	63
4.4.2	入力データの次元について	64
4.4.3	カーネル関数の選定	64
4.4.4	実験結果	65

4.4.5 考察 — グルーヴ感の分布について	75
第5章 結論	77
5.1 まとめ	77
5.2 今後の課題	80
参考文献	82
発表文献	85
謝辞	86

目次

1.1	シーケンサソフト起動画面例 [2]	2
1.2	グローヴ感による物理的特徴パラメータの制御	5
2.1	標準的なドラムの概観	8
2.2	8ビートのドラム譜	10
2.3	16ビートのドラム譜	10
2.4	各楽器・各打拍の打点時刻のメトロノーム音からのずれ(8ビート/100BPM/ ルース) [10]	14
2.5	各楽器・各打拍の打点時刻のメトロノーム音からのずれ(8ビート/100BPM/ タイト) [10]	14
2.6	各楽器・各打拍の打点音量(8ビート/100BPM/ルース) [10]	15
2.7	各楽器・各打拍の打点音量(8ビート/100BPM/タイト) [10]	15
2.8	入力ベクトルが2次元の場合の特徴空間における線形分離超平面(直線)の例	17
2.9	ソフトマージンアルゴリズムにおけるサポートベクター	20
2.10	2次元空間での分類問題における例 [15]	22
3.1	提案手法による実験の全体の流れ	27
3.2	ドラム演奏収録の様子	29
3.3	リズムパターン1	30
3.4	リズムパターン2	30
3.5	リズムパターン3	30
3.6	リズムパターン4	30
3.7	ヒルベルト変換の手順 [18]	34
3.8	ヒルベルト変換による包絡曲線の抽出の例	35

3.9	元の波形	36
3.10	元の波形・ヒルベルト変換により求めた包絡曲線・平滑化後の曲線の比較	36
3.11	ハイハットシンバルの打点時刻と仮想メトロノームの位置との関係	39
3.12	線形分離超平面の係数による重要な特徴の判別	41
4.1	ハイハットシンバル 1 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	49
4.2	ハイハットシンバル 1.5 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	49
4.3	ハイハットシンバル 2 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	50
4.4	ハイハットシンバル 2.5 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	50
4.5	ハイハットシンバル 3 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	51
4.6	ハイハットシンバル 3.5 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	51
4.7	ハイハットシンバル 4 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	52
4.8	ハイハットシンバル 4.5 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	52
4.9	バスドラム 1 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	53
4.10	バスドラム 3 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	53
4.11	スネアドラム 2 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	54
4.12	スネアドラム 4 拍目 分離超平面の方程式における各次元の係数の絶対値の比較	54
4.13	ハイハットシンバル 1 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	69

4.14	ハイハットシンバル 1.5 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	69
4.15	ハイハットシンバル 2 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	70
4.16	ハイハットシンバル 2.5 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	70
4.17	ハイハットシンバル 3 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	71
4.18	ハイハットシンバル 3.5 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	71
4.19	ハイハットシンバル 4 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	72
4.20	ハイハットシンバル 4.5 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	72
4.21	バスドラム 1 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	73
4.22	バスドラム 3 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	73
4.23	スネアドラム 2 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	74
4.24	スネアドラム 4 拍目 打点時刻ずれの標準偏差・打点音量の平均・分離曲線	74

表 目 次

3.1	解析対象の曲（演奏音源データ）一覧	31
3.2	クロスバリデーションの例	40
4.1	各楽器・各打拍における線形 SVM の最適パラメータ	46
4.2	線形 SVM の分離超平面による分類正解率	47
4.3	分離超平面の方程式における各次元の係数・定数項の値	48
4.4	分離超平面の方程式における各次元の係数の絶対値の順位獲得回数	55
4.5	各楽器・各打拍における非線形 SVM の最適パラメータ（4次元特徴空間）	60
4.6	非線形 SVM の分離超曲面による分類正解率	61
4.7	線形 SVM と非線形 SVM の分類正解率の比較	62
4.8	各楽器・各打拍における非線形 SVM の最適パラメータ（2次元特徴空間）	66
4.9	非線形 SVM の分離曲線による分類正解率	67
4.10	分離曲線を表す図の凡例	68

第1章 序論

本章では、まず 1.1 節において、本研究を行うにあたっての背景について述べる。具体的には、1.1.1 節ではコンピュータ音楽について、1.1.2 節では本研究で重要な概念となるグルーブ感の概要と背景について、1.1.3 節ではドラムがグルーブ感に与える影響について、それぞれ述べる。そして、1.2 節において本研究の目的を述べ、1.3 節では本論文の構成を紹介する。

1.1 背景

1.1.1 コンピュータ音楽

近年の音楽業界では、コンピュータを用いた音楽作成手法が盛んに用いられている。この手法では、シーケンサと呼ばれるコンピュータの自動演奏制御ソフトウェアを用いて、楽器の種類・音の高さ・音の長さ・音量などといった演奏制御情報のデータを入力する。

シーケンサソフトを起動させて利用している画面の一例を、図 1.1 に示す。このようにシーケンサ上では、複数のトラックの譜面を一度に管理・編集することができる。また、それぞれのトラックに対して別々に音色や音量等を設定することができる。このデータ入力には、以下のような方法によって行われる。

- マウスを使って画面上の譜面に音符や休符を逐一入力する。
- 鍵盤装置を用いてリアルタイムに入力を行う。
- 音の高さ・音の長さ・音量などの演奏制御情報を数値で直接入力する。

そして、コンピュータに内蔵ないし接続されたシンセサイザ等の音源モジュール機器に収録された音源を用いて、入力した譜面データに基づきコンピュータに自動演奏をさせることができる。

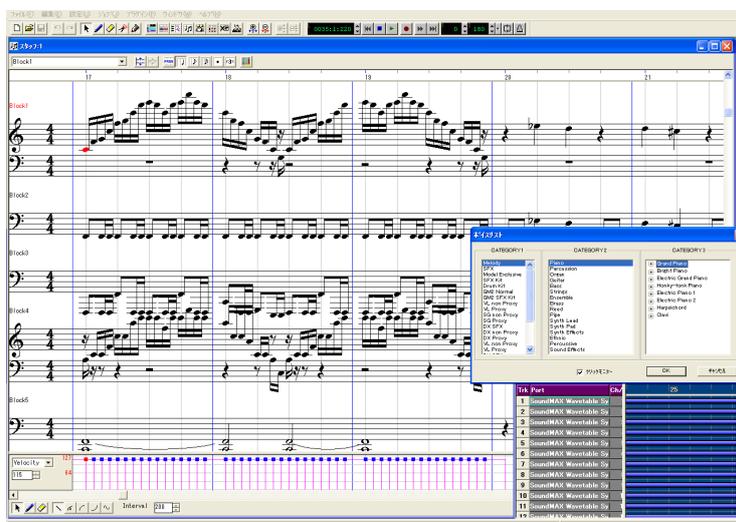


図 1.1: シーケンサソフト起動画面例 [2]

このようにコンピュータを用いて作曲を行う手法は、通称「打ち込み」とも呼ばれる。また、演奏を制御する各種データの規格には、主に MIDI (Musical Instrument Digital Interface) が用いられる。MIDI は統一規格であるため、入出力機器の種類や製造メーカーなどを選ぶことなく、柔軟にメッセージのやりとりが可能である [1]。

MIDI をはじめとする演奏制御情報を用いたコンピュータ音楽技術を用いることによって、従来の作曲・演奏手法と比べて以下のような利点を得ることができる。

- 実際に楽器の演奏ができなくても作曲が可能である。つまり、コンピュータひとつで手軽にオーケストラ編成の音楽すら作成が可能である。
- 楽器が演奏できる場合でも、例えば特定のパートの楽器奏者が身近にいない場合にその楽器パートのみをコンピュータに自動演奏させて代用として使う、といったことが可能である。
- 実際の音の波形ではなくコンピュータ音楽に必要な制御情報のみを用いるため、配信を行う際の通信量を軽量化できる。

このような利点から、ポップス音楽の制作現場を始めカラオケ業界や携帯電話の着信メロディ業界においても、このコンピュータ音楽技術は大変重要な地位を占めている。また、個

人の趣味で音楽制作を行うアマチュアユーザにも広く浸透しており、主に個人的な趣味として行われるコンピュータ上での作曲行為は DTM (DeskTop Music¹) と呼ばれ親しまれている。

以上のように、コンピュータ音楽は今日の音楽業界全体の発展に大きく寄与している技術であり、社会のインフラとしても重要な技術となっている。

1.1.2 コンピュータによるグルーブ感の再現

一方で、コンピュータ音楽では入力された制御数値に演奏が極めて忠実であるがゆえに、文字通りの機械的な演奏が生成されてしまうという問題がある。ある種の楽器を実際に人間が演奏する際は、正確な一定テンポからは若干のずれが生じており、また音程・音量なども微妙に変化する。こうしたずれはグルーブ感とも呼ばれる。このグルーブ感が、リズムに表情を与え機械的でない自然な音楽を生み出している要因のひとつとなっている。

また、グルーブ感にも様々な種類がある。例えば、「カッチリしたリズム」「ゆったりしたリズム」などといったものである。人間が聞き比べれば雰囲気が違うと分かるものの、そうしたグルーブ感の差を生み出す要因が何であるのか、明確には分かっていない。音量の違い・打点時刻の違い・あるいはそれらの分散の違いなど、その要因には様々な可能性が考えられる。グルーブ感同士の境界を定量的に特定することで、グルーブ感の違いをコンピュータが再現することも可能になることが期待できる。

しかし、コンピュータ音楽による音楽制作では、テンポや音程が常に正確無比に制御されている音楽が前提となっている。そのためテンポや音程などのずれが発生せず、結果としてグルーブ感というものが存在しづらい。こうしたグルーブ感の有無に起因する差異が、機械的な音楽を生み出してしまう一因となっている。コンピュータによる音楽作成の熟練者は、このような微妙なずれを数値として意図的に入力することで、実際の人間の演奏に近いグルーブ感を再現している。しかしながら、その労力は膨大である上、自然な演奏を意図的に再現することは困難を極める。

もし、実際の人間の演奏に内在する様々なずれ、つまりグルーブ感を定量的に解析することができれば、その結果をシーケンサソフトウェアのアルゴリズムに応用し、いわゆる「人

¹和製英語であり、DTP (DeskTop Publishing) をもじて使われ始めた言葉とされている。

間らしい演奏」を自動的に行うシステムの構築が可能となる。人間の演奏に関するこれまでの研究は感性的・主観的・定性的なものが大半であり、定量的な解析はあまり行われていない。こうした解析が進めば、コンピュータ音楽に携わる人間の労力が大幅に軽減されることが期待できる。

1.1.3 ドラムによるグルーブ感

ここで、楽曲のグルーブ感を大きくリードしているのはドラムである。例えば、イギリスのロックグループである Led Zeppelin の持つドラムの独特のグルーブ感は「ブリテッシュハードロックの聖域」 [3] といわれるほど独自なものであるとされており、この例のようにドラムのグルーブ感は音楽全体の雰囲気をも大きく支配すると考えることができる。

また、ドラムは他の楽器と比べて音数も少なく単純であり、さらに近年のポピュラー音楽においてはドラム部分をコンピュータで作成している楽曲が非常に多く存在する。以上のような観点から、従来あまり行われてこなかった音楽の定量的解析への入り口として考えても、またグルーブ感の解析という目的と照らし合わせても、ドラム演奏の解析は、グルーブ感の再現に関する問題の解決手段として適している。

1.2 本研究の目的

本研究では、人間のドラム演奏に内在するグルーブ感の定量的な解析を行うことを目的とする。

本研究の目的の概要を、図 1.2 に示す。本研究では、図 1.2 に示すように、打点時刻や音量といった物理的な特徴とグルーブ感の種類との間に写像関係を見出すべく解析を試みる。これにより、ルースやタイトといったグルーブ感 (2.2.2 節参照) の調節によって打点時刻や音量といった物理的特徴パラメータを間接的・自動的に制御できるような環境の実現を目指す。

本研究の位置づけは次の通りである。人間の楽器演奏やグルーブ感に関する従来の定量的解析の研究は各種の測定数値をヒューリスティックに分析していたのに対して、本研究は機械学習の手法を用いてグルーブ感の定量的解析をより客観的かつシステムティックに行うことを試みるものである。そして、本研究の成果によって人間のグルーブ感をコンピュータの

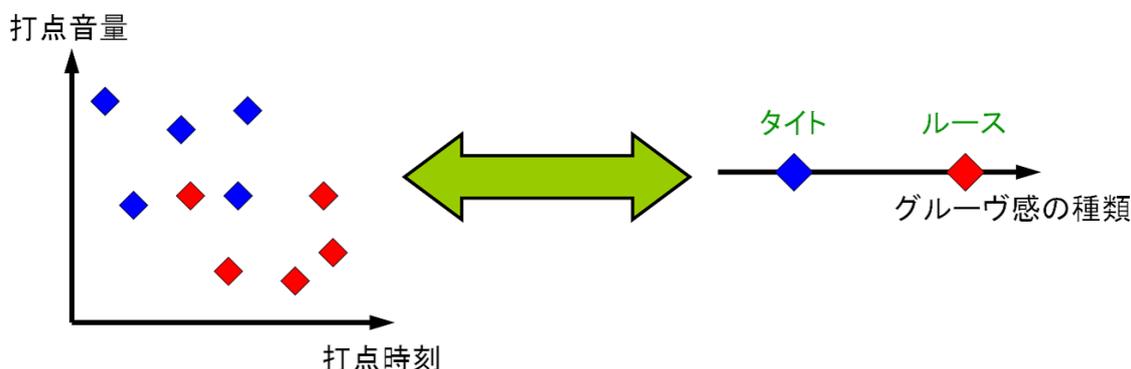


図 1.2: グルーヴ感による物理的特徴パラメータの制御

アルゴリズムとして再現できる環境を実現し、グルーブ感の程度を調整しながら自然な演奏を誰もが気軽に作り出せるような環境の基盤を築き上げることを目指す。

1.3 本論文の構成

以降、本論文の構成は次のようになっている。

第 2 章 関連研究

最初に導入として、ドラムという楽器についての基礎知識の紹介を行う。その上で、音楽の定量的解析に関する研究の紹介、および本研究にて用いる Support Vector Machine (SVM) についてそれぞれ述べる。

第 3 章 提案手法

まず、準備したドラム演奏の演奏音源データについて述べる。その上で、ドラム演奏のグルーブ感を定量的に解析するための本研究の提案手法について説明を行う。

第 4 章 Support Vector Machine によるグルーブ感の解析

まず最初に、実験を行った環境について述べる。そして、本研究にて行った 3 種類の実験について、それぞれの方法と結果を述べる。

第 5 章 結論

本研究の結論と今後の課題について述べる。

第2章 関連研究

本章では、まず導入として2.1節でドラムという楽器に関する基礎知識について述べる。その上で、2.2節において関連研究としてドラム演奏に対して定量的な方法によるアプローチを行っている研究事例について紹介する。次に2.3節において本研究でグローヴ感を分類する際に用いた機械学習の分類器である“Support Vector Machine (SVM)”について説明する。

2.1 ドラムに関する基礎知識

本節では、その詳細についてはあまり一般的に知られていない「ドラム」という楽器について、その基礎的な要素の説明を行う。2.1.1節においてドラムを構成する各打楽器について説明し、次に2.1.2節においてドラムを演奏する際の基本的な演奏パターンについて説明を行う。2.1.3節では、実際に行われているドラム演奏の内容について簡単に紹介する。

2.1.1 ドラムの構成

ドラムとは打楽器のひとつである。ポピュラー音楽などで通常用いられるものは、太鼓やシンバルを複数組み合わせたものであり、「ドラムセット」「ドラムス」などとも呼ばれる。以降、本論文では簡単のためこれを「ドラム」と統一して呼ぶこととする。

標準的な構成のドラムの概観を、図2.1に示す。一般的なドラムを構成する打楽器は、金属筒に張られた膜を叩くことによって音を発する膜鳴楽器群（太鼓類）と、叩くことによってその本体全体が振動して音が出る体鳴楽器群（シンバル類）とに大別される。それぞれの楽器群を構成する打楽器は、以下の通りである。

- 膜鳴楽器群
 - バスドラム (Bass Drum)
 - スネアドラム (Snare Drum)¹
 - フロアトム (Floor Tom)
 - タムタム (Tom-Tom)²
- 体鳴楽器群
 - ハイハットシンバル (Hihat Cymbal)
 - ライドシンバル (Ride Cymbal)
 - クラッシュシンバル (Crash Cymbal)

演奏者は、イスに座った状態でこれらをスティックなどで叩く、およびペダルを踏むことによって演奏を行う。ドラムの演奏で主に用いられる打楽器はハイハットシンバル・バスドラム・スネアドラムの 3 種類であり、この 3 種類の打楽器による打音がドラム演奏の大半を占める。3.3 節にて後述する通り、本研究においてもこの 3 種類の打楽器の演奏を解析の対象とする。

2.1.2 演奏パターン

ドラムの譜面は、通常の楽器の譜面とは異なり、図 2.2 のように表記される。図 2.2 は 8 ビートと呼ばれるリズムパターンの楽譜である。このリズムパターンは、

- 8 分音符のリズムでハイハットシンバルを叩き続ける
- 各小節の 1 拍目と 3 拍目にバスドラムのペダルを踏む
- 各小節の 2 拍目と 4 拍目にスネアドラムを叩く

¹厳密には、スネアドラムは膜を叩くことによって、反対側の膜に張られた響き線が振動して、スネアドラム独特の音が出るので、他の膜鳴楽器とは少々趣きが異なる。なお、本来はこの響き線のことを「スネア (Snare) 」と呼ぶ。

²タムタムは、音の高さによってさらに「ハイトム (High Tom) 」「ロートム (Low Tom) 」に分類される。



図 2.1: 標準的なドラムの概観

というものであり、ドラムを演奏する上で欠かすことのできない基本的かつ重要なリズムパターンのひとつである。このとき、バスドラムを叩く 1 拍目と 3 拍目を強拍、スネアドラムを叩く 2 拍目と 4 拍目を弱拍とも言う。

また、ハイハットシンバルでリズムを刻む間隔を 8 ビートの半分にあたる 16 分音符のリズムにした 16 ビートと呼ばれる基本リズムパターンも存在する。16 ビートの楽譜を図 2.3 に示す。

2.1.3 ドラム演奏の実際

実際のポピュラー音楽では、バスドラムやスネアドラムのパターンを、これらの基本リズムパターンから変化させたり、聞こえるか聞こえないか程度の小さな音量でスネアドラムを叩く「ゴーストノート」と呼ばれるものを付加させたりして演奏されることもある。

また、演奏する曲の雰囲気によっては、ハイハットシンバルの代わりにライドシンバルを叩くことによって 8 分音符や 16 分音符のリズムを刻むこともある。

フロアタム・タムタム・クラッシュシンバルなどは、フィルインの際に多く使われる。フィルインとは、曲の間や変わり目³に入れる短いフレーズの事である。このフィルインを入れることにより、曲をより一層と盛り上げる効果が生まれる。

なお、昨今では上記のような従来のドラムの他に、エレクトリックドラム（電子ドラム）も用いられることがある。エレクトリックドラムとは、通常のドラムと同様に組まれた樹脂製のパッドをスティックなどで叩くことにより、パッドに直結されたシンセサイザ音源からスピーカを介して音を出すドラムである。パッドからの入力をデータ入力としてシーケンサで制御し、コンピュータ音楽に応用することも可能である。ただし、このデータ入力の方法は当然ながらドラム演奏技術を持っている人間でないと不可能である。

³例えば、A メロから B メロへのつなぎ部分、あるいは B メロからサビへのつなぎ部分などのことを指す。

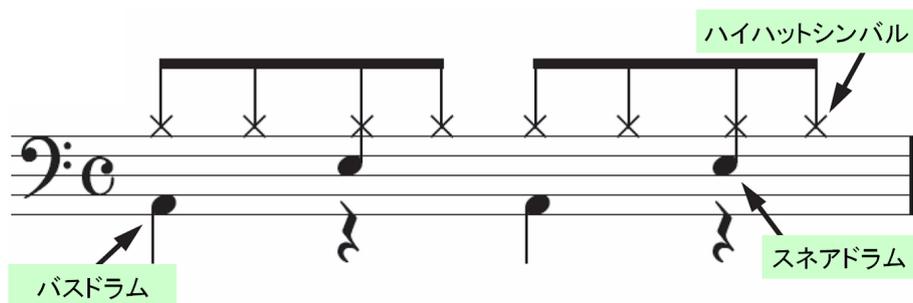


図 2.2: 8 ビートのドラム譜

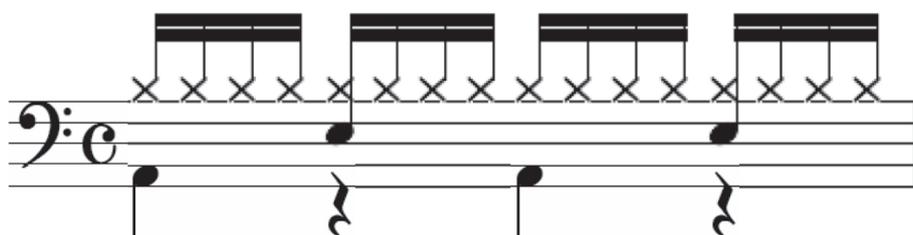


図 2.3: 16 ビートのドラム譜

2.2 人間のドラム演奏に関する研究

ドラムをはじめとする楽器を人間が演奏する際には、正確なテンポからのずれや音量の微妙な変化が生じる。すなわち、これが演奏に潜むグルーヴ感である。しかし、音楽分野の従来の研究における人間の演奏に対する考察は、主観的あるいは感性的なアプローチによるものが大半を占める。従って、人間の演奏に対する定量的・科学的な解析はあまり進んでいないのが現状である。以下では、人間の楽器演奏—とりわけドラム演奏を対象として定量的なアプローチが行われている数少ない研究例を取り上げる。

2.2.1 ドラム演奏に対する定量的解析

Friberg ら [4] は、ジャズドラム奏者のレコード演奏からシンバルの打点時刻を抽出し、スウィング比との関連についての研究を行っている。スウィングというのは、シンバル等でリズムを刻むときの連続する 2 つの打点の時間間隔を変化させる演奏手法のことを指す。スウィングの無い通常の 8 分音符が 1 拍を 2 等分するのに対し、スウィングを効かせた演奏の場合は 1 拍を 2:1 や 3:1 といった変則的な比 (= スウィング比) に分けて 8 分音符を演奏することがある。Friberg らの研究においては、

- 演奏する曲のテンポが遅い際にはスウィング比が約 3.5:1 であった
- 曲のテンポが速くなってゆくにつれてスウィング比が 1:1 に近づいていった

などといった成果が示されている。人間の演奏が機械的ではなく柔軟に変化するものであることが定量的に示されている例である。

また、同じくジャズの分野において、Waadeland ら [5] はジャズドラム奏者の手の動きとシンバル打点時刻の関連を調査し、運動学的な観点からの議論を行っている。

辻ら [6] は、モーションキャプチャシステムを用いて、ドラム奏者の演奏情報（打拍時刻、打拍の強さ）とフォーム情報（手の甲の動き）の同時測定を行った。その結果、打楽器経験者と未経験者とは、リズム安定性や叩く強さの安定性などについて定量的に有意な差があったとの結論を見出している。

2.2.2 ドラム演奏のグルーブ感に関する研究

ここでは、ドラム演奏に関する研究の中でも特に「グルーブ感」に注目したものを取り上げる。

河瀬・中村・長岡・Draguna ら [7, 8, 9] は、熟達したドラムの演奏者が自身の演奏を聴取した際に感じるグルーブ感について心理学的な観点からの調査を行った。その結果、ドラム演奏者が演奏時に意図し表現したグルーブ感を、聴取時にも自身で判別できることを示した。また、人工演奏音による聴取実験を行うことにより、楽譜からのずれあるいはテンポのばらつきの特徴が聴取においてグルーブ感に影響を及ぼし、これらの要因がグルーブ感に関係があることなどを示した。

奥平・平田・片寄ら [10, 11, 12, 13] は、ポップス系音楽においてドラムのグルーブ感は、その違いにより楽曲全体の印象を変えるような重要な要素のひとつであると捉え、グルーブ感の定量的解析を行った。この研究では、主に 8 ビートに分類される 4 種類のリズムパターンについて、プロのドラム奏者による 2 種類のグルーブ感の演奏を収録し、その音データからハイハットシンバル・バスドラム・スネアドラムの打点時刻および音量を測定し、解析を行っている。

ここで収録を行った 2 種類のグルーブ感というのは、一般に「ルース」「タイト」と呼ばれるものである。ルースの聴感上の印象はゆったりした滑らかなグルーブ感であり、一方タイトの聴感上の印象はきちんとまとまった明解なグルーブ感であると言える。このような聴感上の印象は多くのドラム奏者の間で共通しており、暗黙知の一例と見なすこともできる。

ドラム奏者は、ヘッドフォンから流されるメトロノーム音（一定テンポで流れるクリック音⁴）を聞きながら、それに合わせて演奏を行った。演奏は楽器別に収録され、得られたサンプリングデータから各楽器のメトロノーム音からの時刻の差分および音量を求め、統計処理を行った。例として、2.1.2 節の図 2.2 で示した 8 ビートのリズムを 100BPM⁵のテンポで演奏した際の音源について、各楽器の発音時刻を検出しメトロノーム音からの時刻の差分を

⁴ ドンカマとも呼ばれる。

⁵ Beat Per Minute : 1 分間あたりの拍数を表す指標。例えば、マーチ等に採用される 1 秒に 2 拍のテンポなら、120BPM となる。

解析した結果を、図 2.4 および図 2.5 に示す。図 2.4 はルースなグルーブ感で演奏したもので、図 2.5 はタイトなグルーブ感で演奏したものである。図中、横軸は拍を、縦軸は各打点の時間変化分の 16 分音符に対する割合を百分率で示したものを表す。○, ×, △ はそれぞれハイハットシンバル、バスドラム、スネアドラムを表している。図 2.4 および図 2.5 を比較すると、ルースな演奏（図 2.4）の方はタイトな演奏（図 2.5）の場合よりも打点時刻が遅れる傾向があり、またそのばらつきも大きくなる傾向があることがわかる。

同様に、同演奏の音量について各楽器ごとに音量を算出し統計処理を行った結果の一例を、図 2.6 および図 2.7 に示す。図 2.6 はルースなグルーブ感で演奏したもので、図 2.7 はタイトなグルーブ感で演奏したものである。図中、横軸は拍を、縦軸は各楽器の収録演奏中に最大音量を 100%としたときの各打点の音量を百分率で示したものを表す。○, ×, △ はそれぞれハイハットシンバル、バスドラム、スネアドラムを表している。図 2.6 および図 2.7 を比較すると、ルースな演奏（図 2.6）の方はタイトな演奏（図 2.7）の場合よりも音量が小さくなる傾向があるが、ばらつきはそれほど大差ないことがわかる。

この研究では、これらの図をはじめとして他にも様々な切り口からの統計処理が行われた。そこから導き出されたとされている結論の一部をここで示すと、以下のようなものである。

- ルースの演奏はタイトの演奏に対し、打点時刻が遅れる傾向が見られた。
- ルースの打点時刻の分散がタイトの打点時刻の分散よりも大きいものが多かった。
- テンポが速くなると、打点時刻の遅れが大きな値になる。

またこの研究では、さらにゴーストノートを付加させたドラムパターンについても同様の研究が行われた。そして、以上のような解析結果をもとにした、ドラム演奏生成システムの実装も行われた。

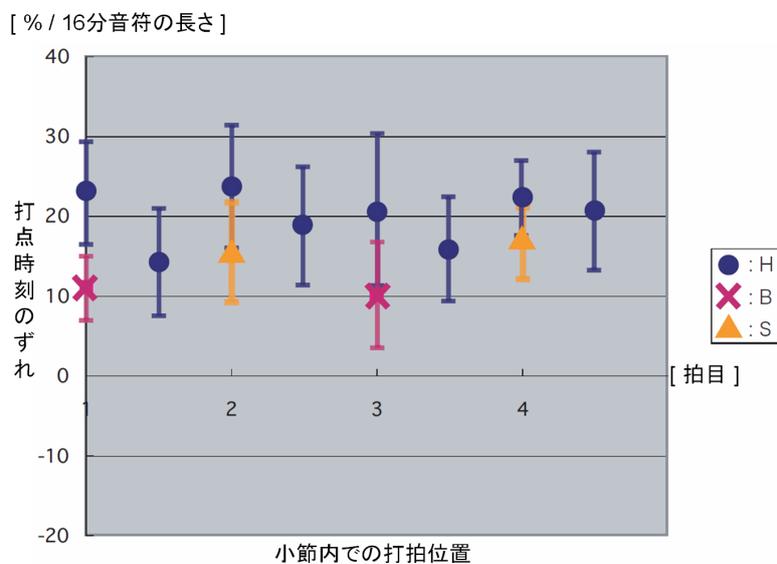


図 2.4: 各楽器・各打拍の打点時刻のメトロノーム音からのずれ (8 ビート/100BPM/ルース) [10]

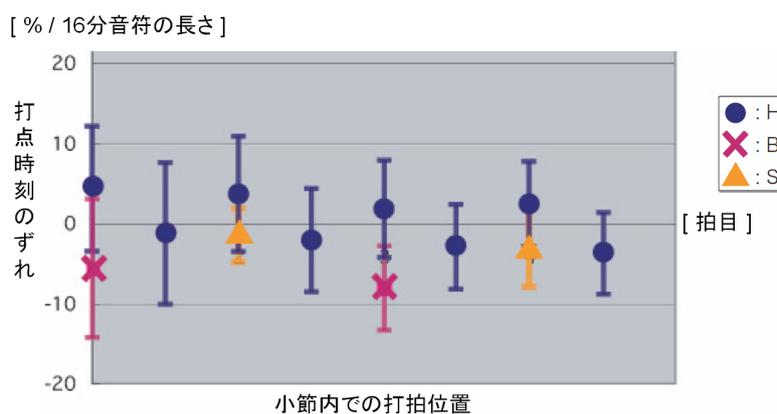


図 2.5: 各楽器・各打拍の打点時刻のメトロノーム音からのずれ (8 ビート/100BPM/タイト) [10]

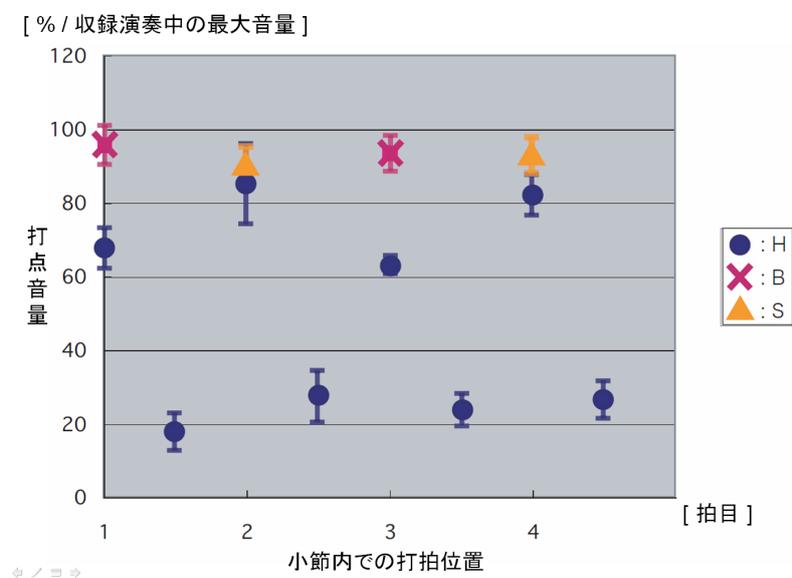


図 2.6: 各楽器・各打拍の打点音量 (8 ビート/100BPM/ルース) [10]

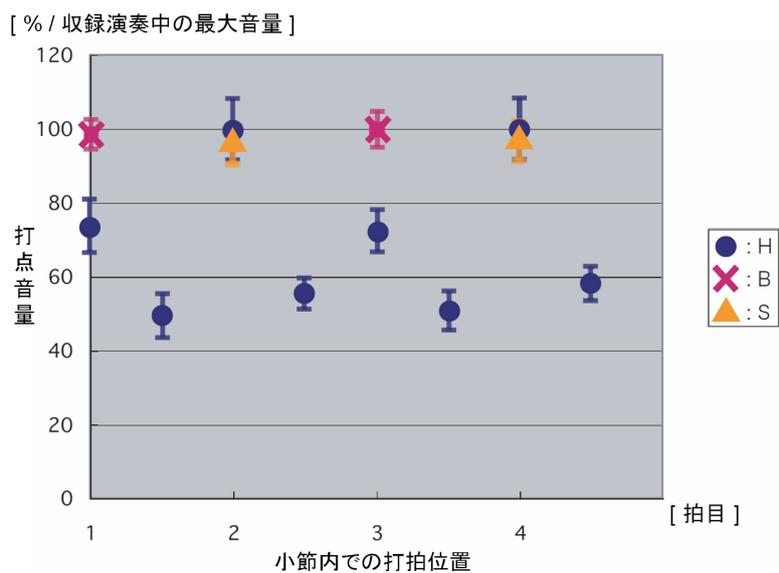


図 2.7: 各楽器・各打拍の打点音量 (8 ビート/100BPM/タイト) [10]

2.3 Support Vector Machine

Support Vector Machine (SVM) は、1992 年に Vapnik ら [14] によって提案された、2 クラスの分類器のひとつである。SVM とは、超平面とそれに最も近い例の距離（マージン）を最大化するように、訓練データを線形分類する分離超平面を求める手法である。

以下では、まず基本となる線形 SVM による分類問題の解法について述べる。その上で、より多くの問題に SVM を柔軟に対応させるための手法として、分類の精度を緩和するソフトマージナルゴリズム、および非線形の分類への拡張を可能とするカーネルトリックについて述べる。

2.3.1 線形 SVM による分類問題の解法

2 値クラス分類では、実数値関数 $f : X \subseteq \mathbb{R}^n \mapsto \mathbb{R}$ を用いて、入力ベクトル $x = (x_1, \dots, x_n)^T$ から $f(x) \geq 0$ の値を求め、その正負によって割り当てるクラスを識別する。この際に用いられる実数値関数 $f(x)$ は、 $x \in X$ に関する線形関数の場合を考えると、

$$f(x) = w^T x + b \quad (2.1)$$

と書くことができる。これを幾何的に解釈すれば、方程式 $f(x) = 0$ は「入力空間 X を 2 つに分類する超平面」を表す。ベクトル w は分離超平面の直交する方向を定義し、 b の値を変化させることで分離超平面は平行移動する。2 次元入力ベクトルによる特徴空間における分離超平面の例を、図 2.8 に示す。図 2.8 は、 印のクラスと  印のクラスを分類した例である。

SVM では、入力ベクトルを完全に識別する超平面の中で、2 つのクラスの「真ん中」を通るものを最も優れた分離超平面と考える。従って、SVM では、超平面と入力ベクトルとの最小距離を最大化するように分離超平面を決定する。このときに、分離超平面との最小距離に相当する入力ベクトル、つまり分離超平面に最も近い入力ベクトルをサポートベクター⁶と呼ぶ。図 2.8 において、内部が色で塗りつぶされている点がサポートベクターである。一般に、サポートベクターは各クラスに必ずしも 1 個ずつとは限らない。

ここで、各入力ベクトル x_1, \dots, x_n に対し、それぞれのクラスラベル y_1, \dots, y_n を設定し、入力ベクトル x_i がクラス A に所属していれば $y_i = -1$ 、クラス B なら $y_i = 1$ とする。

⁶日本語訳の慣習に基づき、ここでは「サポートベクトル」ではなく「サポートベクター」と呼ぶこととする。

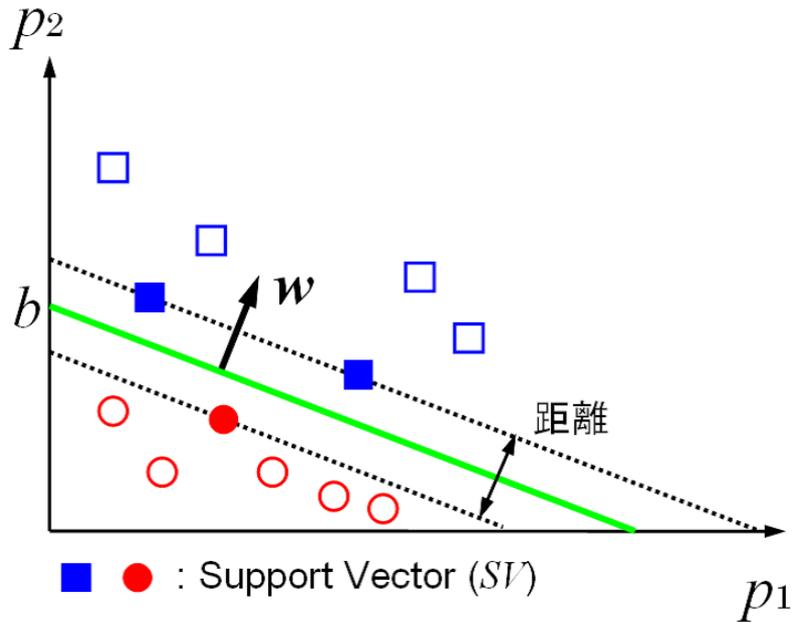


図 2.8: 入力ベクトルが 2 次元の場合の特徴空間における線形分離超平面 (直線) の例

パラメータ w, b は, 定数倍しても表現する超平面が全く変わらないため, 冗長性を持っている. この冗長性を解決し学習結果を一意に定めるため, 以下の制約を加える.

$$\min_{i=1, \dots, n} |w^T x_i + b| = 1 \tag{2.2}$$

この制約を加えることにより, 入力ベクトルと分離超平面の最短距離

$$\min_{i=1, \dots, n} \frac{|w^T x_i + b|}{\|w\|} \tag{2.3}$$

は, 常に $\frac{1}{\|w\|}$ となる.

w および b は, 入力ベクトルを完全識別するものの中から, 最小距離を最大にするように決めるので, 次のように定式化される.

$$\min_w \|w\|^2 \tag{2.4}$$

$$\text{制約条件 } y_i (w^T x + b) \geq 1 \tag{2.5}$$

目的関数を最小化すれば，最小距離 $\frac{1}{\|\mathbf{w}\|}$ の最大化となる．また，制約条件は，この超平面が入力ベクトルを完全識別することを示している．

以下，ラグランジュ未定乗数法を用いると，未定乗数 $\alpha_i (\geq 0)$ を導入し，ラグランジュ関数

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i \{y_i (\mathbf{w}^T \mathbf{x}_i + b)\} \quad (2.6)$$

を得る．これを，条件 $\frac{\partial L}{\partial b} = \frac{\partial L}{\partial \mathbf{w}} = 0$ のもとで解くと，次のような関係が導かれる．

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (2.7)$$

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad (2.8)$$

これを，式 (2.4)(2.5) に代入すると， α のみに関する最大化問題となる．

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (2.9)$$

$$\text{制約条件 } \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (2.10)$$

このとき，Karush-Kuhn-Tucker 条件

$$\alpha_i (y_i \mathbf{w}^T \mathbf{x}_i - 1) = 0 \quad (2.11)$$

が要求されるため，分離超平面から一番近い入力ベクトル，すなわちサポートベクター以外の入力ベクトルに対応する α は全て 0 になる．サポートベクターの集合を SV とすると，式 (2.9) より求められる最適解 α^* を用いて， w の最適解 w^* およびそのときの b の値は，

$$\mathbf{w}^* = \sum_{\mathbf{x}_i \in SV} \alpha_i^* y_i \mathbf{x}_i \quad (2.12)$$

$$b = y_k - \mathbf{w}^{*T} \mathbf{x}_k \quad (\mathbf{x}_k \in SV) \quad (2.13)$$

となる．

以上より，クラス識別に用いる実数値関数 $f(\mathbf{x})$ は，

$$f(\mathbf{x}) = \mathbf{w}^{*T} \mathbf{x} + b = \sum_{\mathbf{x}_i \in SV} \alpha_i^* y_i \mathbf{x}_i^T \mathbf{x} + b \quad (2.14)$$

となり，求める分離超平面の方程式は

$$\sum_{\mathbf{x}_i \in SV} \alpha_i^* y_i \mathbf{x}_i^T \mathbf{x} + b = 0 \quad (2.15)$$

と求められる．

2.3.2 ソフトマージンアルゴリズム

2.3.1 節にて述べた手法は，対象となるデータ群が分離超平面によって完全に識別できることを仮定としている．しかし，一般に多くの問題は線形分類不可能であることが多い．その場合，式 (2.5) を満たす w, b が存在しないため，最適化を行うことが不可能となる．

このような場合に用いられる手法の一つとして，ソフトマージンアルゴリズムが挙げられる．ソフトマージンアルゴリズムとは，多少の識別誤りを許すように制約を緩める手法である．ここでは，代表的なソフトマージンアルゴリズムであり本研究においても用いる C-SVC [14] について述べる．

制約条件を緩めるために，以下のスラック変数

$$\xi_i \geq 0, \quad i = 1, 2, \dots, n \quad (2.16)$$

を導入する．そして，式 (2.4) にあった最適化の目的関数および式 (2.5) にあった最適化の際の制約条件を次のように変更する．

$$\min_w \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (2.17)$$

$$\text{制約条件} \quad y_i (\mathbf{w}^T \mathbf{x} + b) \geq 1 - \xi_i \quad (2.18)$$

ここで，式 (2.17) 中に登場する変数 C は識別誤りに対するペナルティであり， C の値が大きくなる程，識別誤りを許さない方向に最適化が行われる．なお， C は経験的・実験的に決定されるパラメータであり，汎用的に C を決定する手法は現在のところ確立されていない．

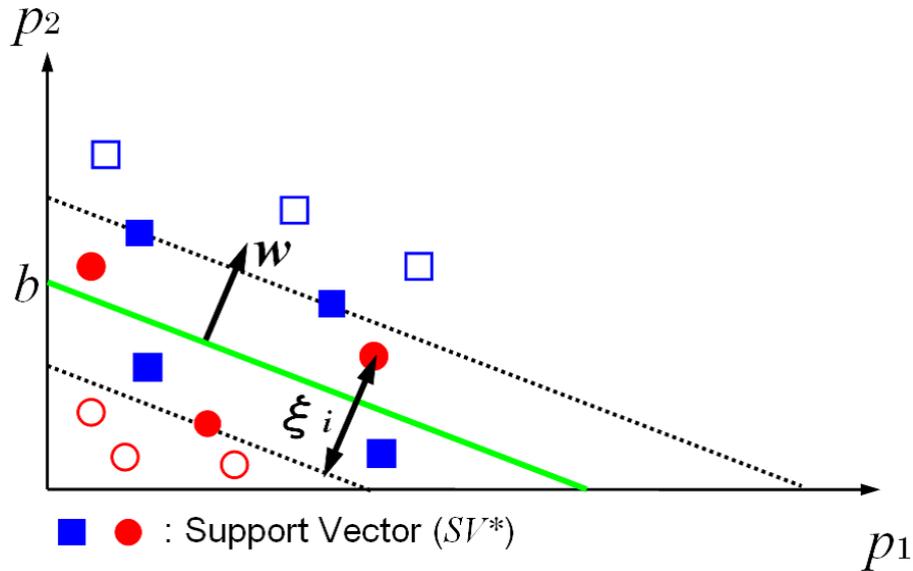


図 2.9: ソフトマージナルゴリズムにおけるサポートベクター

このように最適化問題を変更することにより，ラグランジュ未定乗数 α_i に関する式 (2.9)・式 (2.10) は以下のように変更される．

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (2.19)$$

$$\text{制約条件} \quad 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (2.20)$$

ソフトマージナルゴリズムにおいては， $\alpha_i > 0$ であるような点を全てサポートベクターと呼ぶ⁷．ソフトマージナルゴリズムを用いた場合のサポートベクターおよびスラック変数 ξ_i の様子を，図 2.9 に示す．図 2.9 のように，ソフトマージナルゴリズム適用後のサポートベクターの集合を SV^* とすると，式 (2.19) は式 (2.9) と同様に解くことができ，ソフトマージナルゴリズムを用いた場合の分離超平面の方程式は式 (2.15) と同様の以下の

⁷流儀によっては，あくまで分離超平面に最も近い入力ベクトルのみをサポートベクターと呼ぶ場合もあるが，本論文においては， $\alpha_i > 0$ であるような点 — つまり分離超平面を求める計算の際に考慮するベクトルを全てサポートベクターと呼ぶこととする．

式となる .

$$\sum_{\mathbf{x}_i \in SV^*} \alpha_i^* y_i \mathbf{x}_i^T \mathbf{x} + b = 0 \quad (2.21)$$

この分離超平面の形状は、先に述べたパラメータ C の値によって変化するため、分類性能は C の値に大きく依存する。従って、本研究においても 3.7 節にて述べるように適切な C の値の設定を行うものとする。

2.3.3 カーネルトリック — 非線形分類への拡張

2.3.2 節にて述べたソフトマージナルゴリズムによって、分類可能な問題の数は増大する。しかし、複雑な現実的アプリケーションによっては、線形分類だけでは分類することができない。そこで、さらに表現力のある非線形分類を用いることが不可欠となってくる。非線形分類に際して主に用いられる手法が、カーネルトリックと呼ばれる手法である。以下、カーネルトリックについてその概要を簡単に述べる。

カーネルトリックは、線形分類が不可能なデータ群を非線形写像によって高次元の特徴空間へ写像し、高次元空間で線形モデルを用いて分類を行うことにより、元の空間で非線形分類を行っていることと実質的に等価になるという手法である。

以下の写像 $\phi(\mathbf{x})$ を考える。

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \mapsto \phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \dots, \phi_N(\mathbf{x})) \quad (2.22)$$

$\phi: \mathbb{R}^n \mapsto \mathbb{R}^N$ は N 次元空間への非線形写像であり、 N は元の次元数 n よりも大きいものとする。高次元空間への写像を行うことにより、元の空間では線形分類が不可能な問題であっても高次元空間において線形分類が可能となる。その例を図 2.10 に示す。この例では、2 次元空間に左図のようにサンプルが分布している場合、直線（1 次元の線形モデル）によるクラス分類を行うことは不可能である。しかし、以下の写像

$$\begin{aligned} \phi: \mathbb{R}^2 &\mapsto \mathbb{R}^3 \\ (x_1, x_2) &\mapsto (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2) \end{aligned} \quad (2.23)$$

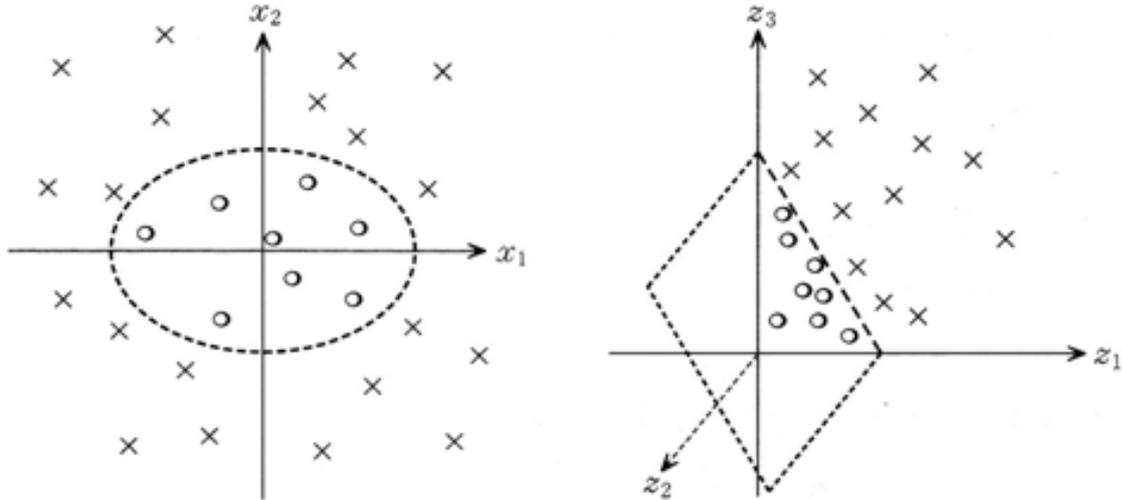


図 2.10: 2次元空間での分類問題における例 [15]

を用いてこれを 3次元空間へ写像すると，図 2.10 の右図のように平面（2次元の線形モデル）によるクラス分類が可能となる．

ここで， ϕ に対応する以下のような関数 $K(x, z)$ を考える． x, z は任意のベクトルとする．

$$K(x, z) = \phi(x)^T \phi(z) \quad (2.24)$$

この関数 K をカーネル関数と呼ぶ．カーネル関数を用いることにより，高次元空間での線形モデルは K のみによって表現することができ，分離超平面の式 (2.15) は

$$\sum_{\mathbf{x}_i \in SV} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b = 0 \quad (2.25)$$

と書き換えることができる．これにより，実際には写像 ϕ の計算を回避することができる．この仕組みをカーネルトリックという．カーネルトリックの利点は，非常に高次元の空間における内積 $\phi(x)^T \phi(z)$ の計算が入力空間のカーネル関数のみによって行えるため，次元の呪いとも呼ばれる高次元空間特有の計算量爆発を抑えつつ非線形分類問題を解くことが可能であるという点である．代表的なカーネル関数には，以下のようなものが挙げられる．

- 多項式カーネル

$$K(x, z) = (\gamma x^T z + r)^d, \quad \gamma > 0 \quad (2.26)$$

- Radial Basis Function (RBF) カーネル

$$K(\mathbf{x}, \mathbf{z}) = \exp\left(-\gamma \|\mathbf{x} - \mathbf{z}\|^2\right), \quad \gamma > 0 \quad (2.27)$$

- シグモイドカーネル

$$K(\mathbf{x}, \mathbf{z}) = \tanh\left(\gamma \mathbf{x}^T \mathbf{z} + r\right)^d \quad (2.28)$$

ここで、各式中に登場する γ, r, d はそれぞれカーネル関数のパラメータである。2.3.2 節にて述べたソフトマージンアルゴリズムにおけるパラメータ C と同様に、カーネル関数のパラメータの値も SVM の分類性能に大きく寄与するものであるため、本研究では、これらのパラメータの値についても 3.7 節にて述べるように適切な値の設定を行うものとする。

2.3.4 SVM の利点

SVM の主な特徴として、以下のようなものを挙げるができる。

- マージン最大化基準を採用した識別手法である。
- 凸 2 次計画問題を解くことにより、最適な識別関数が一意に定まる。
- ソフトマージンアルゴリズムを用いることにより、幅広い問題への適用が可能である。
- 線形構造を持たない場合であっても、カーネルトリックを用いて非線形分類を行うことができる。

このような特徴を持つことにより、SVM には汎化能力が非常に高いという利点がある。つまり、与えられるデータ集合がどのようなものであっても、比較的高い正答率のクラス分類が可能である。また、同じ機械学習手法であるニューラルネットワークのように局所解問題に陥る心配が少ないことも、SVM の利点である。このような利点から、SVM は昨今の多くの応用研究に活用されている。

第3章 提案手法

本章では、まず本研究のアプローチを 3.1 節で述べた上で、提案手法による実験全体の概要を 3.2 節で述べる。次に、実験の内容に関する詳細な説明として、3.3 節では解析の対象であるドラムの演奏音源に関する説明を行う。3.4 節では本研究において SVM を用いる根拠を述べる。そして以下、データを解析するために本実験にて提案する各種手法として、3.5 節では音源から打点位置を検出する手法、3.6 節では打点時刻のずれを求める基準となる仮想メトロノームの位置を決定する手法、3.7 節では SVM のパラメータを決定する手法、3.8 節では SVM による実験結果の解釈の方法に関する説明を、それぞれ行う。

3.1 本研究のアプローチ

本節では、2 章にて述べた関連研究を踏まえた上で、グローヴ感の解析に対する本研究のアプローチについて述べる。

2.2.2 節で紹介した研究においては、ルースとタイトという 2 種類のグローヴ感について、それらの違いに関する定量的な考察が行われている。しかしその際には、得られた平均や分散といった数値の大小を研究者自身が見て比較をするなど、グローヴ感の解析をヒューリスティックに行っている箇所が多く見受けられる。

また、従来の研究においては、標準偏差ゼロの場合にドラム音を鳴らす予定の基準時刻やドラム演奏者がヘッドフォンで聞くメトロノーム音の時刻など、コンピュータによって正確に刻まれる一定のテンポを基準とすることで各楽器の打点時刻のずれを定量評価している。しかし現実問題として、演奏者が聞いているようなこうしたメトロノーム音を、ドラム演奏を実際に聞く立場の人間が聞くことは当然できない。

これらに対し本研究では、以下のようなアプローチをとる。まず従来研究における解析手法にヒューリスティックな点が多く存在することに対しては、より客観的・定量的かつ汎用性の高いシステマティックな解析手法を確立すべく、本研究では機械学習を用いた解析を行う。打点に関する各種データ値を機械学習で解析することにより、値の単純な大小にとられない総合的な観点からルースとタイトの差異に関する考察を進める。楽器演奏の解析に機械学習の手法を用いた研究には前例がなく、機械学習の手法の汎用性を示すという意味でもこのアプローチは有意義であると言える。本研究では、機械学習の手法の中でも特に汎化能力の高い Support Vector Machine (SVM) を用いる。数ある機械学習の手法の中から SVM を採用するの背景については 3.4 節にて詳しく述べる。

また、演奏を聞く立場の人間（ユーザ）がメトロノームの音を聞くことができない点に対しては、本研究ではユーザが実際に聞くことのできる演奏音のみを扱うことによって、よりユーザに近い立場でのグルーブ感の解析を試みる。具体的には、演奏の聴取者が脳内で補完するであろう仮想的なメトロノーム音（以降、仮想メトロノームと呼ぶ）の時刻を求め、その仮想メトロノームからの各楽器の打点時刻のずれを算出し、各種検証を行ってゆくものとする。なお、この仮想メトロノームに関する詳細な説明は 3.6 節にて行う。

3.2 提案手法の概要

提案手法による実験の全体の流れを図 3.1 に示す。本実験では、まず解析対象となるドラム演奏を収録する。本研究の目的はグルーブ感の違いを解析することであるので、ドラム演奏はルースなグルーブ感とタイトなグルーブ感の両方をプロのドラム奏者に演奏してもらい、収録を行う。

収録した演奏データから、楽器の叩かれた時刻と音量を抽出する。この際、信号解析の手段としてヒルベルト変換・平滑化・極大値検出という 3 つの段階を経ることとする。また、抽出したハイハットシンバルの打点時刻から、演奏を聞く人間が脳内で補完するであろう仮想メトロノームの位置を算出し、仮想メトロノームと実際の打点時刻との差分（打点時刻ずれ）を求める。

打点時刻ずれ・打点音量を抽出したら、そこから各曲における打点時刻ずれ・打点音量の平均値・標準偏差をそれぞれ求める。これら計 4 種類の特徴の値を入力値とする 4 次元特徴

空間上において、まず線形 SVM によってグルーブ感（ルース・タイト）を分類する分離超平面を生成する。生成された分離超平面の形状から、打点時刻ずれ・打点音量の平均値・標準偏差のうちどの特徴がグルーブ感の分類に大きく寄与しているのか判別を行う。

次に、同じ 4 次元特徴空間において非線形 SVM によってグルーブ感を分類する分離超曲面を生成する。この分類結果と、線形 SVM の場合の分類結果との比較を行うことにより、グルーブ感の分類問題における両者（線形 SVM・非線形 SVM）の分類性能の差について議論を行う。

また、グルーブ感の分類に寄与していると判別された上位 2 種の特徴について、さらに 2 次元特徴空間において非線形 SVM による分離曲線を生成する。そして、分類正解率を 4 次元特徴空間の場合と比較するとともに、グルーブ感の具体的な分類状況を図示し 2 次元特徴空間においてグルーブ感の分布状況の解析を行う。

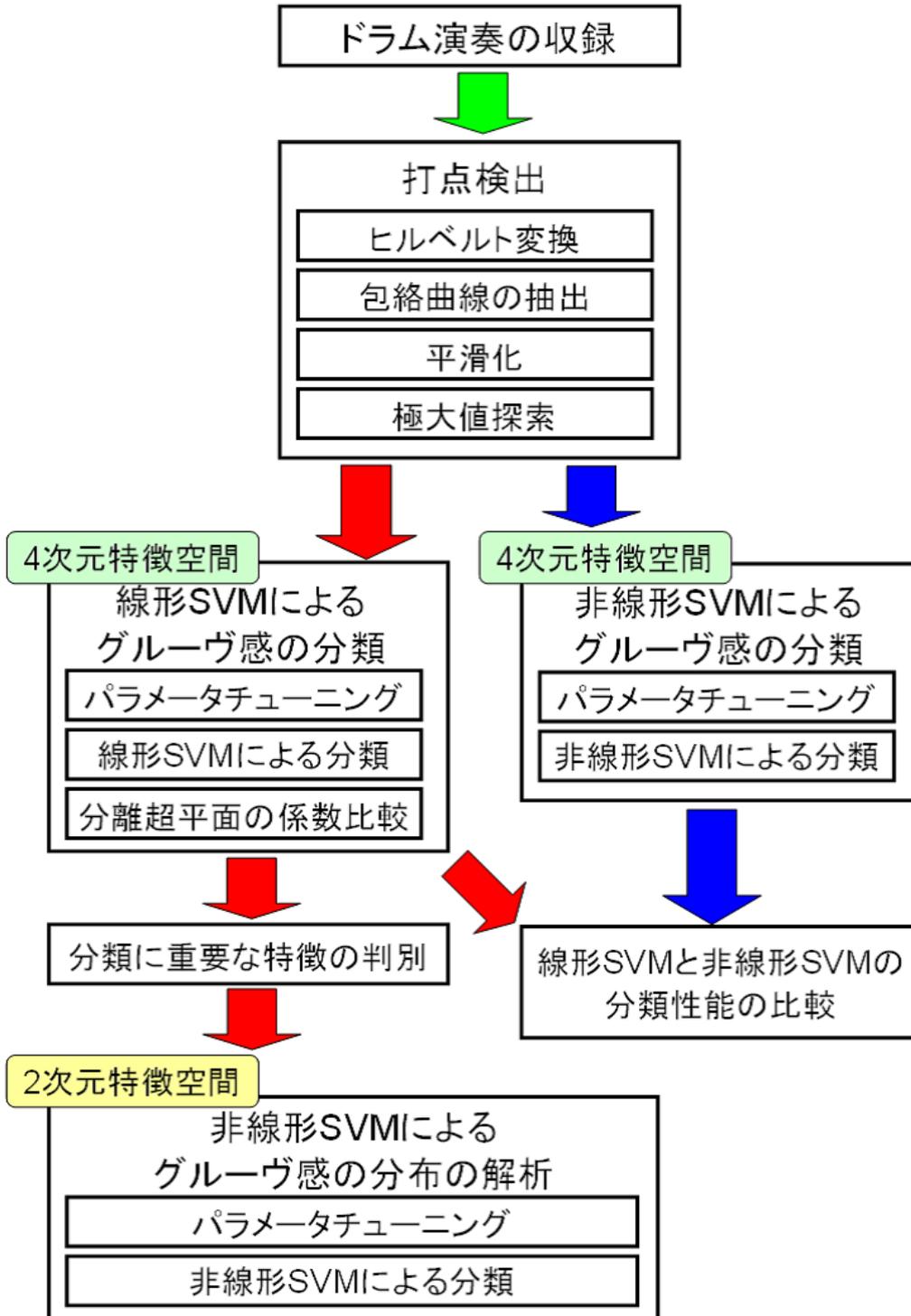


図 3.1: 提案手法による実験の全体の流れ

3.3 解析対象の音源データ

本研究では、2.2.2 節で述べた奥平・平田・片寄ら [10, 11, 12, 13] による研究で用いられているものと同一の音源を解析対象とする。この音源は、プロのドラム奏者に演奏を依頼し、数種類のリズムパターンおよびテンポに対し、ルースとタイトの 2 種類のグルーブ感でドラム演奏を行ったものを収録した音源である。

演奏収録は、Digidesign 社の Pro Tools¹ [17] を用い、ハイハットシンバル・バスドラム・スネアドラムに対して、それぞれ個別にマイクを配置して各打楽器の音を別トラックに分けて収録された。すべての演奏収録を通して、打楽器配置、マイク位置、Pro Tools のボリューム設定などの録音状態は一定に保った。

ドラム奏者に対しては次のように演奏するように求めた。

- ドラム奏者は全ての演奏において、右手でハイハットシンバル・左手でスネアドラムをそれぞれ叩き、右足でバスドラムのペダルを踏むくことによって演奏を行う。
- ドラム奏者は、一定テンポで流れるメトロノーム音をヘッドフォンで聴きながら、それに合わせて演奏する。

演奏収録の様子を図 3.2 に示す。以上により、各打楽器ごとに別々に収録された 44.1 KHz・16bit の WAVE データが得られた。

今回収録したドラム演奏については、2 種類のグルーブ感の違いの他にテンポやリズムパターンなど様々な組み合わせの指示をドラム奏者に与えて演奏を収録している。このうち、本実験では以下の演奏音源データについて解析を行う。

- テンポスピード … 2 パターン (100 BPM / 132 BPM)
- リズムパターン … 4 パターン (いずれも 8 ビートを基本としたもの)

リズムパターンについて、今回用いた 4 つのリズムパターンを、図 3.3～図 3.6 にそれぞれ示す。図 3.3 のリズムパターン 1 を基準として、図 3.4 のリズムパターン 2 は小節の 2.5 拍

¹オーディオの録音・編集・ミキシングなど音楽制作上必要となる一連の作業を総合的に行うことのできるデジタルオーディオワークステーション (DAW) のひとつ。各種エフェクトのプラグインが充実しており、音楽制作現場をはじめとして映画・放送など多くの分野においてデファクトスタンダードとしての地位を確立している。



図 3.2: ドラム演奏収録の様子

目・図 3.5 のリズムパターン 3 は小節の 3.5 拍目・図 3.6 のリズムパターン 4 は小節の 2・4 拍目にそれぞれバスドラムを追加で鳴らすリズムパターンとなっている。これらでは、バスドラムが余分に入ることにより他の楽器の演奏にも影響及ぼす。

各演奏音源データは、それぞれ 8 小節のドラム演奏から成る。便宜上、以下ではこの 8 小節のドラム演奏の音源データを曲と呼び、音源データ 1 種類につき 1 曲と数えることとする。

今回解析を行う音源データは計 17 曲である。その内訳を表 3.1 に示す。なお、表 3.1 の曲番号 13 と 14 については、同じ条件の曲を 2 回収録してある。これらの曲を解析し打点位置を検出することで、各曲・各楽器・各打拍における以下の 4 種類の値を求める。

- 打点時刻ずれの平均
- 打点音量の平均
- 打点時刻ずれの標準偏差
- 打点音量の標準偏差

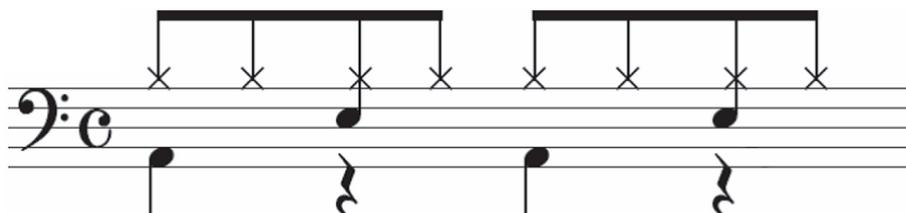


図 3.3: リズムパターン 1

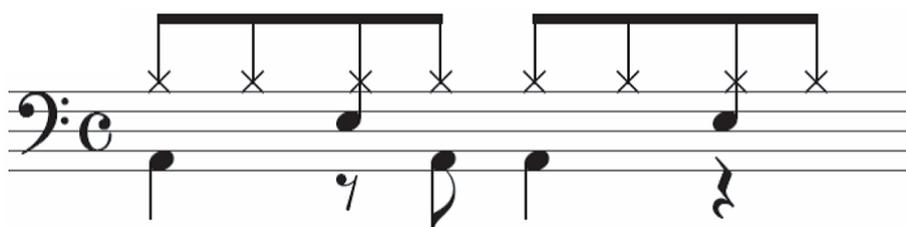


図 3.4: リズムパターン 2

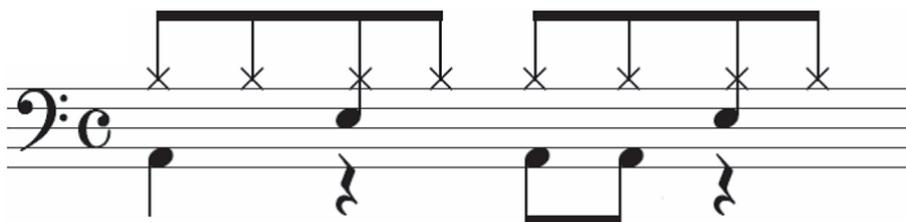


図 3.5: リズムパターン 3

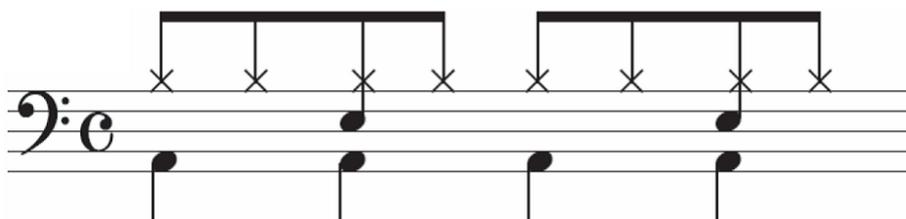


図 3.6: リズムパターン 4

表 3.1: 解析対象の曲 (演奏音源データ) 一覧

曲番号	リズムパターン	テンポ (BPM)	グルーヴ感
1	パターン 1	100	ルース
2	パターン 1	100	タイト
3	パターン 1	132	ルース
4	パターン 1	132	タイト
5	パターン 2	100	ルース
6	パターン 2	100	タイト
7	パターン 2	132	ルース
8	パターン 2	132	タイト
9	パターン 3	100	ルース
10	パターン 3	100	タイト
11	パターン 3	132	ルース
12	パターン 3	132	タイト
13	パターン 4	100	ルース
14	パターン 4	100	ルース
15	パターン 4	100	タイト
16	パターン 4	132	ルース
17	パターン 4	132	タイト

ここで、打点時刻ずれというのは、実際の打点時刻と仮想メトロノームとの時刻の差のことである。本論文では、以降でも打点時刻ずれという言葉はこの意味で用いる。

3.4 SVM 使用の背景

ここでは、本実験において何故 SVM を用いるかについて説明する。これに対しては、2.3.4 節に述べたように、SVM は汎化能力に優れた分類器であるということが理由として挙げられる。

本実験において対象とするデータは人間のドラム演奏である。しかし、2.2 節にて述べたように、人間の演奏に関して定量的な解析を行った研究は例が少なく、データの特徴がよくわかっていない。よって、データをグルーブ感によって分類するためには、どのようなデータに対しても分類能力を発揮できる分類器、つまり汎化能力に優れた分類器が必要不可欠である。

そこで本実験では、汎化能力に優れた分類器である SVM を分類器として採用し、ルースとタイトの 2 クラス分類を試みる。

3.5 打点検出方法

本研究では、収録した演奏音源の波形を分析し、打点を検出して打点時刻と音量を抽出する。ここでの手順は、大きく分けて次の 3 つに分けられる。

1. ヒルベルト変換による包絡曲線の抽出
2. 包絡曲線の平滑化处理
3. 極大値探索による打点の検出

以下では、これらの 3 つの手順についてその手法の詳細をそれぞれ述べる。

3.5.1 ヒルベルト変換による包絡曲線の抽出

打点検出にあたっては、基本的には波形の振幅が極大となる部分を打点として検出する方法が考えられる。そしてこれを実現する手段として、単純に各時刻における WAVE データ

値の絶対値を振幅として、それが極大となる場所を探し出すという手法が考えられる。しかしこの手法では、実際には真の振幅が大きい位相の関係で変位が小さくなっているという箇所であっても「振幅が小さい」と判定されてしまうため、正しい振幅を把握できなかったことにはならない。あくまで各時刻の WAVE データ値は位相の影響を含む変位に過ぎず、各時刻における瞬時振幅は別に存在すると考えなければならない。

ここで、瞬時振幅を求める手法であるヒルベルト変換という処理について説明する。ヒルベルト変換は以下の手順 [18] によって行われる。

1. 元の波形データを高速フーリエ変換 (FFT) する。
2. FFT によって得られたスペクトルのうち、周波数 0~0.5 の成分を 2 倍に (ただし周波数 0 の DC 成分は例外でそのままにして)、周波数 0.5~1 の成分をゼロにする。
3. 手順 2 により求めたスペクトルを逆高速フーリエ変換 (IFFT) する。得られた複素数信号 $x_a(n)$ (解析信号という) の実部を $r(n)$ 、虚部を $h(n)$ とすると、虚部 $h(n)$ がヒルベルト変換となる。

このヒルベルト変換の手順を図 3.7 に示す。このとき、任意の時刻における瞬時振幅 $A(n)$ は、

$$A(n) = \sqrt{r(n)^2 + h(n)^2} \quad (3.1)$$

という式で算出できる。これは、ヒルベルト変換を施した信号が元の信号に対して全ての周波数帯域において位相が $\pi/2$ 遅れた信号になるという性質を利用している。そして、この瞬時振幅を繋いでいったものが、元の波形データの包絡曲線となる [19]。ヒルベルト変換による包絡曲線の抽出の例を、図 3.8 に示す。この図では、元の波形が黒い曲線・ヒルベルト変換によって抽出された包絡曲線が赤線でそれぞれ表示されている。ヒルベルト変換により確かに包絡曲線が抽出されていることが確認できる。

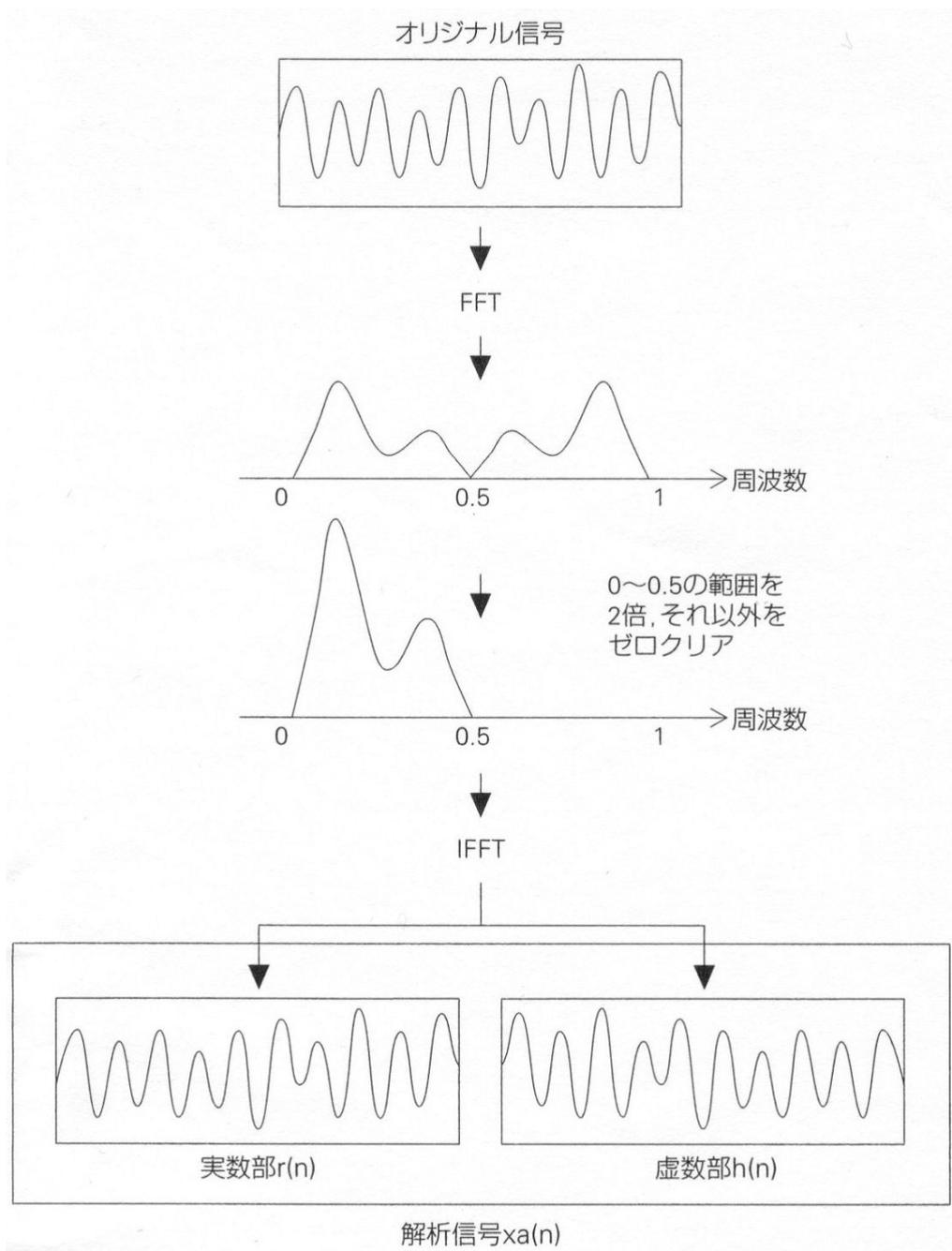


図 3.7: ヒルベルト変換の手順 [18]

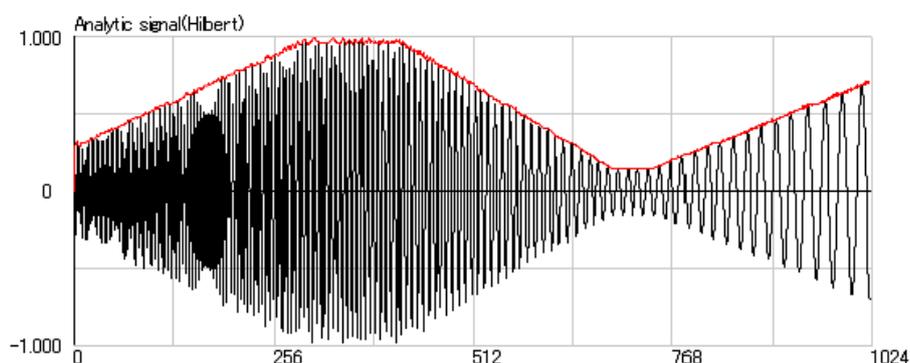


図 3.8: ヒルベルト変換による包絡曲線の抽出の例

3.5.2 包絡曲線の平滑化処理

一般に、打楽器の音の波形は非常に複雑な形状をしている。そのため、そのままでは極大値が無数に存在してしまうため波形のどの箇所が発音部分に相当するのを探し出すことは難しい。

そこで以下では、包絡曲線の極大値の探索を簡単にするべく、包絡曲線の平滑化を行う。本実験では、以下の手順にて平滑化を行う。

1. 初期データとして、3.5.1 節のヒルベルト変換によって求めた包絡曲線のデータを与える。
2. 与えられたデータについて、そのデータ自身の値と前後 n 個ずつの計 $2n + 1$ 個のデータの値の平均値を計算し、得られた値をあらたにそのデータの値とする。
3. 手順 2 の平滑化処理を計 k 回繰り返して行う。

本実験では、各楽器の波形の立ち上がり形状などを考慮し、上記手順中の各パラメータの値については $n = 250$, $k = 3$ として平滑化処理を行った。ここで、本実験で収録した音源データの一部を取り出した元の波形を図 3.9 に示し、図 3.9 の元の波形にヒルベルト変換により求めた包絡曲線と平滑化後の曲線のデータ値を重ねて比較したものを図 3.10 に示す。

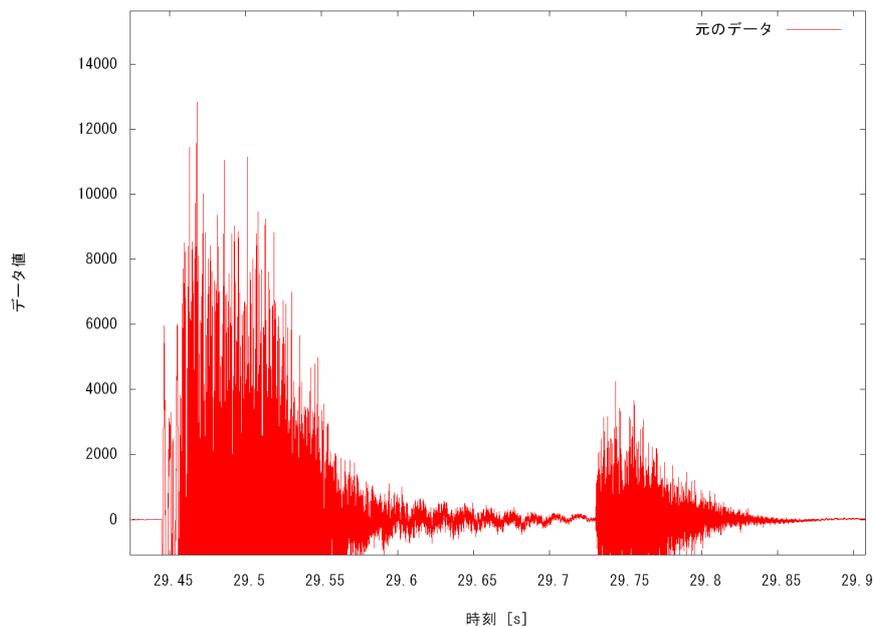


図 3.9: 元の波形

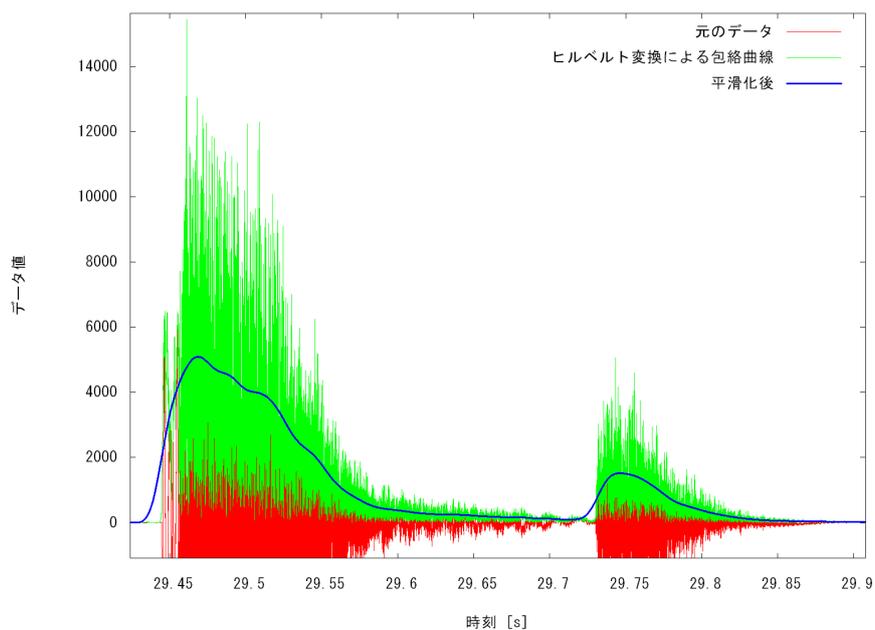


図 3.10: 元の波形・ヒルベルト変換により求めた包絡曲線・平滑化後の曲線の比較

3.5.3 極大値探索による打点の検出

平滑化を施したデータに対して、以下の手順で極大値探索による打点検出を行う。

1. 幅 w サンプル分の探索窓を用意し、これを平滑化後のデータサンプル列の先頭から時間軸方向に進めてゆく。
2. 探索窓の右側（時間軸の正方向側）にあたるサンプルの振幅が、探索窓の左側（時間軸の負方向側）の振幅よりも一定値 h だけ高くなったら、打点開始フラグを立てる。
3. 打点開始フラグが立った状態で、探索窓の右側の振幅が左側の振幅以下となったら、その時点での探索窓の内部に打点があると判断し、探索窓の移動を一時停止する。
4. 探索窓内部で振幅が最大となる点を探索し、得られた最大地点の振幅を打点の音量、時間軸座標を打点時刻とする。
5. 手順 4 の打点検出が終了したら、再び探索窓の移動を再開する。
6. 探索窓がデータサンプル列の末尾に到達するまで、同様にして打点検出を繰り返す。

本実験では、各楽器の波形の立ち上がり形状などを考慮し、上記手順中の各パラメータの値を、 $w = 1000$, $h = 100$ として解析を行った。

この結果、各演奏データに対して打点時刻を検出することに成功した。また、検出した打点時刻に相当する箇所における平滑化後データの値を打点音量とした。

3.6 仮想メトロノーム

3.1 節にて述べたように、本研究ではドラム演奏者が聞くメトロノーム音は使わずに、ユーザが実際に聞くことのできる演奏音のみを扱う。そこで本実験では、演奏音中に含まれるハイハットシンバルのリズムから、演奏の聴取者が脳内で補完するであろう仮想メトロノームの時刻を求める。そして、その仮想メトロノームからの各楽器の打点時刻のずれを算出し、各種検証を行ってゆくものとする。以下、仮想メトロノームの間隔と開始時刻を求め位置を決定する手法について述べる。

一回の演奏の中で、ハイハットシンバルが連続して計 N 回叩かれているとき、その打点時刻を t_1, t_2, \dots, t_N とする。また、求める仮想メトロノームの開始時刻を t 、リズムを刻む時間間隔を d とする。これらの関係を、図 3.11 に示す。

このとき、ハイハットシンバルの打点時刻と仮想メトロノームの時刻との差分の二乗値の合計は、以下の 2 変数関数

$$m(t, d) = \sum_{k=1}^N [t_k - \{t + (k-1)d\}]^2 \quad (3.2)$$

によって表されるので、この関数 m を最小化するような t と d を求めることにより、仮想メトロノームの位置を決定することができる。ここで、関数 m を最小化する t と d は、以下の 2 元連立方程式

$$\frac{\partial m}{\partial t} = 0 \quad (3.3)$$

$$\frac{\partial m}{\partial d} = 0 \quad (3.4)$$

を解くことによって求められ、それぞれ、

$$t = \frac{4(N+1) \sum_{k=1}^N t_k - 6 \sum_{k=1}^N kt_k}{N(N+1)} \quad (3.5)$$

$$d = \frac{12 \sum_{k=1}^N kt_k - 6(N+1) \sum_{k=1}^N t_k}{N(N+1)(N-1)} \quad (3.6)$$

となる²。

以下の実験では、式 (3.5)・式 (3.6) 中に現れる $\sum_{k=1}^N t_k$ および $\sum_{k=1}^N kt_k$ は実際の演奏音源中のハイハットシンバルの打点時刻を求めることによって計算を行う。そして、その結果を式

²式の導出の過程で、級数に関する以下の公式

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$$

を用いた。

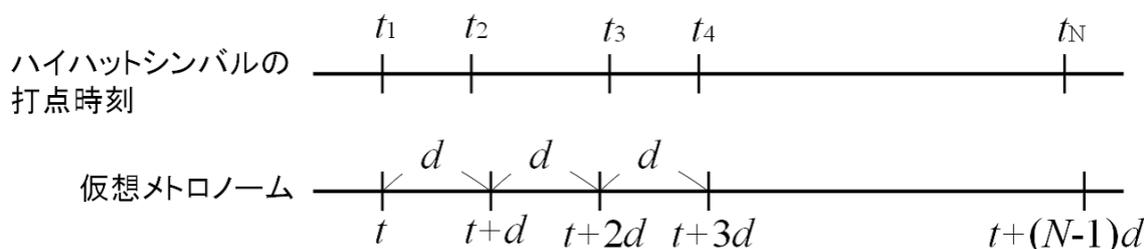


図 3.11: ハイハットシンバルの打点時刻と仮想メトロノームの位置との関係

(3.5)・式 (3.6) に代入することによって t と d の値を求め、仮想メトロノームの位置を決定する。仮想メトロノームの位置が決まったら、仮想メトロノームと実際の打点時刻との差を打点時刻ずれとして記録する。

3.7 クロスバリデーションによる SVM パラメータチューニング

SVM を用いるに際しては、2.3.2 節で述べたソフトマージンアルゴリズムのパラメータ C や、2.3.3 節で述べたカーネル関数のパラメータを設定する必要があるが、これらのパラメータの値は SVM の挙動に大きな影響を与えるものであるため、根拠ある適切な値の設定を行う必要がある。例えば、 C の値を非常に大きなものにして分離超平面を生成すれば、与えられたデータ群に対しては非常に正解率の高い分離超平面となるが、しかしこれでは過学習 (overfitting) となり他の新たな入力に対する汎用性が無くなってしまいう可能性がある。本実験に即して言うと、 C を大きくして分離超平面を生成すれば本実験のために収録を行ったドラム演奏に対してのグルーヴ感の分類正解率は高くなるものの、将来的に新たな演奏のデータを投入した際に、その分離超平面ではグルーヴ感の適切な分類ができなくなる恐れがあるということである。

こうした過学習を避け適切なパラメータの値を設定するために、本実験ではクロスバリデーション (cross validation) を用いたパラメータチューニングを行う。クロスバリデーションは学習モデルの汎化性を調査するための方法のひとつであり、交差検定法とも呼ばれる。標準的な n 分割クロスバリデーション (n -fold cross validation) の手順は以下の通りである。

1. モデル構築に必要なパラメータの値を設定する (SVM であれば、ソフトマージン

表 3.2: クロスバリデーションの例

手順 (順序は任意)	学習データ	テストデータ	分類の成否
1	A, B, C	D	正解
2	A, B, D	C	正解
3	A, C, D	B	不正解
4	B, C, D	A	正解
平均分類正解率 = 75%			

ルゴリズムのパラメータ C やカーネル関数のパラメータを設定する)。

2. 対象データ群全体を n セットに分割する。
3. $n - 1$ 個のセット (学習データ) を用いてモデルの構築を行う (SVM であれば, 分離超平面を生成する)。
4. 構築したモデルを残った 1 セット (テストデータ) に適用し, 分類精度の評価を行う (SVM であれば, 残りの 1 セットが分離超平面により正しくクラス分類されたかどうかを評価する)。
5. 手順 3・4 の操作を, 全てのデータがテストデータとして用いられるように繰り返す。
6. 全ての評価値の平均を取り, 学習モデルの性能を評価する (SVM であれば, 最初に設定したパラメータの値がどれほど適切であったかの評価を行う)。

例として, 対象データが A,B,C,D の 4 個である場合に 4 分割クロスバリデーションを行うときの手順を表 3.2 に示す。このように, 各手順においてテストデータの分類が正解か不正解かを記録し, クロスバリデーション全体での平均分類正解率を算出する。この平均分類正解率を用いて, 学習モデルの性能の評価, すなわち各種パラメータの値の設定の妥当性の評価を行う。

本実験では, ソフトマージナルゴリズムのパラメータ C やカーネル関数のパラメータを様々に変化させながら, それぞれのパラメータの組み合わせに対してクロスバリデーションによる分離超平面の評価を行い, もっとも評価の高かったパラメータの値と, そのパラ

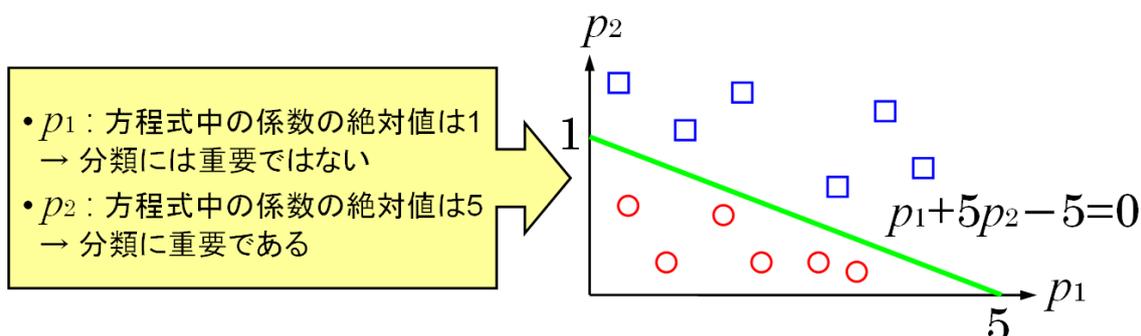


図 3.12: 線形分離超平面の係数による重要な特徴の判別

メータによって生成される分離超平面を採用することとする。なお、3.3 節で述べたように本実験では 17 曲分のデータを用いるため、各パラメータの組み合わせに対して 17 分割クロスバリデーションを行うことによりパラメータ設定の評価を行う。

3.8 線形分離超平面の係数による重要な特徴の判別

本節では、SVM による実験結果の解釈の方法として、SVM により生成した線形分離超平面の係数を用いる方法について述べる。

本実験では、まずは SVM によりルースとタイトの線形分類を試みる。そして、SVM によって生成された分離超平面の形状を調べることにより、入力ベクトルのどの次元が分類に大きく関与しているかを判別し、ルースとタイトというグルーヴ感の違いを生み出す要因となっている特徴の判別を行う。

ここで、線形分離超平面の形状から分類要因の判別を行うプロセスは以下の通りである。例として、図 3.12 を参照いただきたい。この図中では、線形分離超平面（この場合は 2 次元の特徴空間なので、線形分離超平面は直線となる）の形状が、 p_1 軸に対して平行に近い傾きになっている。すなわち、この図の場合は、 \square のクラスと \circ のクラスを分類している大きな要因となっている特徴は p_2 であり、 p_1 は 2 クラスの分類にそれほど関与していないことがわかる。

一般に、平面の方程式に現れる各変数の係数は法線ベクトルの成分を表すので、SVM によって求められた分離超平面の方程式の中での係数の絶対値が大きいほど、分離超平面がその次

元を明確に分類していることになる．図 3.12 中にある分離超平面の方程式は $p_1 + 5p_2 - 5 = 0$ となっており，係数の絶対値はそれぞれ p_1 が 1， p_2 が 5 となっている．上記の通り，クラスを分類している大きな要因は特徴 p_2 である．よって，方程式中の各次元の係数の絶対値の比較により分類の要因となっている特徴を特定することが確かに可能となっている．

このようにして，本実験においても，ルースとタイトを分ける線形の分離超平面の各次元の係数をもとにして，ルース・タイトのグローヴ感を分類している主な要因の特徴の判別を行う．

第4章 SVMによるグルーブ感の解析

本章では、3章にて述べた提案手法を用いて行った実験について述べる。まず、4.1節では本実験のデータ処理を行った実験環境について述べる。その上で、4.2節では、実際にドラム演奏のグルーブ感を線形SVMで解析した実験の方法と結果について示す。4.3節では、4.2節と同様のデータ条件で非線形SVMによる解析を行い、4.2節での線形SVMによる分類実験結果との比較を行う。さらに4.4節では、解析する次元を絞り非線形SVMでより詳細な解析を行った実験の方法と結果について示す。

4.1 実験環境

本実験のデータ処理は全て、以下のスペックのコンピュータ上にて行った。

- CPU : Intel Pentium 4 — 3.40GHz × 2
- RAM : 2.00GB
- OS : Microsoft Windows XP Service Pack 2

解析の際には、プログラミング言語としてC++およびPythonを用いた。SVMの処理については、SVM専用ライブラリであるLIBSVM [20]をベースにして改良したプログラムを用いた。

4.2 線形 SVM による 4 次元特徴空間での分類と重要な特徴の判別

ここでは、ルース・タイトの両グルーブ感が含まれた 17 曲分のデータに対して線形 SVM による解析を行い、各楽器・各打拍における分離超平面を生成する。そして、生成された分離超平面の方程式における各次元の係数をもとにして、グルーブ感の分類に寄与している主な要因の特徴の判別を行う。

4.2.1 実験手順

ここでの実験の手順は以下の通りである。

1. SVM に用いるデータの準備
2. クロスバリデーションによる線形 SVM のパラメータチューニング
3. 線形 SVM による分離超平面の生成
4. 生成された分離超平面の方程式における各次元の係数の比較

また、SVM による分離超平面の生成は、次の各楽器・各打拍ごとに行う。

- ハイハットシンバル

- 1 拍目
- 1.5 拍目
- 2 拍目
- 2.5 拍目
- 3 拍目
- 3.5 拍目
- 4 拍目
- 4.5 拍目

- バスドラム
 - 1 拍目
 - 3 拍目
- スネアドラム
 - 2 拍目
 - 4 拍目

なお，上記の「 拍目」という表記は，各小節内での拍の位置を示す．

また，データを SVM に入力する際に設定するクラスラベル y_i は，グルーブ感によって以下のような値に設定する．

- ルースなグルーブ感の演奏から得たデータ： $y_i = -1$
- タイトなグルーブ感の演奏から得たデータ： $y_i = 1$

4.2.2 入力データの次元について

ここで解析を行うデータは，各曲ごとに 1 データずつ，計 17 個のデータとする．各データは，各曲における以下の 4 種の特徴の値を持つ 4 次元ベクトルとする．

- 打点時刻ずれの平均
- 打点音量の平均
- 打点時刻ずれの標準偏差
- 打点音量の標準偏差

4.2.3 実験結果

4.2.3.1 クロスバリデーションによるパラメータチューニングの結果

まず，グルーブ感の分類に先立ち，過学習を避けるべくクロスバリデーションによる線形 SVM のパラメータの設定を行った．ここでは，2.3.2 節で述べたソフトマージナルゴリズ

表 4.1: 各楽器・各打拍における線形 SVM の最適パラメータ

楽器	小節内の打拍位置	C	クロスバリデーションの 平均分類正解率 (%)
ハイハットシンバル	1 拍目	0.5	76.4706
	1.5 拍目	512.0	82.3529
	2 拍目	0.125	70.5882
	2.5 拍目	0.5	88.2353
	3 拍目	16.0	100.0000
	3.5 拍目	2.0	94.1176
	4 拍目	8.0	70.5882
バスドラム	4.5 拍目	4.0	88.2353
	1 拍目	2.0	88.2353
	3 拍目	0.25	88.2353
スネアドラム	2 拍目	4.0	82.3529
	4 拍目	512.0	76.4706

ムのパラメータ C について、その値を $2^{-5}, 2^{-4}, \dots, 2^{15}$ と等比数列的に変化させながら、各 C の値を用いて 17 分割クロスバリデーションを逐一行い、クロスバリデーションの平均分類正解率が最も高くなった C の値を採用した。

クロスバリデーションを用いて設定された各楽器・各打拍における C の値と、その際のクロスバリデーションのそれぞれの平均分類正解率を、表 4.1 に示す。以下の線形分類実験では、ここで設定された各楽器・各打拍における C の値を最適パラメータとして用いるものとする。

4.2.3.2 分離超平面の生成と分類正解率

表 4.1 の C の値を用いた線形 SVM によって、ルースとタイトのグルーブ感を分類する分離超平面を、各楽器・各打拍ごとに生成した。なお、ここでは SVM による処理の都合上、入力データの各次元の値は、各楽器・各打拍ごとに $[-1, 1]$ の範囲に正規化されたものを用いている。

まず、生成された分離超平面によるグルーブ感の分類正解率を、表 4.2 に示す。表 4.2 を

表 4.2: 線形 SVM の分離超平面による分類正解率

楽器	小節内の打拍位置	分離超平面による 分類正解率 (%)
ハイハットシンバル	1 拍目	82.3529
	1.5 拍目	100.0000
	2 拍目	76.4706
	2.5 拍目	94.1176
	3 拍目	100.0000
	3.5 拍目	94.1176
	4 拍目	88.2353
バスドラム	1 拍目	88.2353
	3 拍目	88.2353
スネアドラム	2 拍目	82.3529
	4 拍目	100.0000

見ると、分類正解率が最も低かったのはハイハットシンバル 2 拍目の 76.4706% で、最も高かったのはハイハットシンバル 1.5 拍目・ハイハットシンバル 3 拍目・スネアドラム 4 拍目のいずれも 100.0000% であった。また、その他の楽器・打拍についても、軒並み 80% 台後半～90% 台の高い正解率となっている。このことから、ハイハットシンバル・バスドラム・スネアドラムのどの楽器についても、グルーブ感によって演奏に明らかな差が生じているということが確認できたと同時に、線形分類であっても高い正解率での分類が可能であることが示された。

4.2.3.3 分離超平面の係数比較

次に、各楽器・各打拍における分離超平面の方程式中に現れる各次元の係数および定数項の値を、表 4.3 に示す。ただし、分離超平面の方程式 $f(x)$ は

$$f(x) = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + b = 0 \quad (4.1)$$

であり、 x_1, x_2, x_3, x_4 の各次元は、それぞれ以下の特徴に対応しているものとする。

表 4.3: 分離超平面の方程式における各次元の係数・定数項の値

楽器	小節内の打拍位置	a_1	a_2	a_3	a_4	b
ハイハットシンバル	1 拍目	-0.4379	-1.2721	0.2892	0.4440	-0.2892
	1.5 拍目	-7.3730	-22.3474	25.6994	-0.0244	-1.6728
	2 拍目	-0.1434	-0.3420	0.6253	0.5826	-0.0708
	2.5 拍目	0.2476	-0.8413	1.2883	0.6664	-0.2601
	3 拍目	0.7195	-3.7472	4.1272	1.0184	-2.8756
	3.5 拍目	-1.3156	-1.7463	1.0883	0.7699	0.1078
	4 拍目	-0.1106	-0.6649	3.7733	-1.1612	-1.1827
バスドラム	1 拍目	-2.0616	0.8125	0.7273	0.7666	-0.0832
	3 拍目	-0.3267	-0.0192	1.0971	0.3093	-0.0639
スネアドラム	2 拍目	-2.1500	0.3907	1.4855	-1.0190	0.3718
	4 拍目	0.2693	-3.2189	40.4558	-1.9122	-24.0406

- x_1 : 打点時刻ずれの平均
- x_2 : 打点音量の平均
- x_3 : 打点時刻ずれの標準偏差
- x_4 : 打点音量の標準偏差

3.8 節で述べたように、係数の絶対値を比較することによって、上記 4 次元が表す特徴のうち特にどの特徴がグルーヴ感の分類に寄与しているのかを判別することができる。ここで、各楽器・各打拍において分離超平面の方程式における各次元の係数の絶対値を比較したグラフを、図 4.1 ~ 図 4.12 に示す。図 4.1 ~ 図 4.12 において、グラフの縦軸は分離超平面の方程式における各次元の係数の絶対値、グラフの横軸は左から順に打点時刻ずれの平均・打点音量の平均・打点時刻ずれの標準偏差・打点音量の標準偏差の次元をそれぞれ表している。

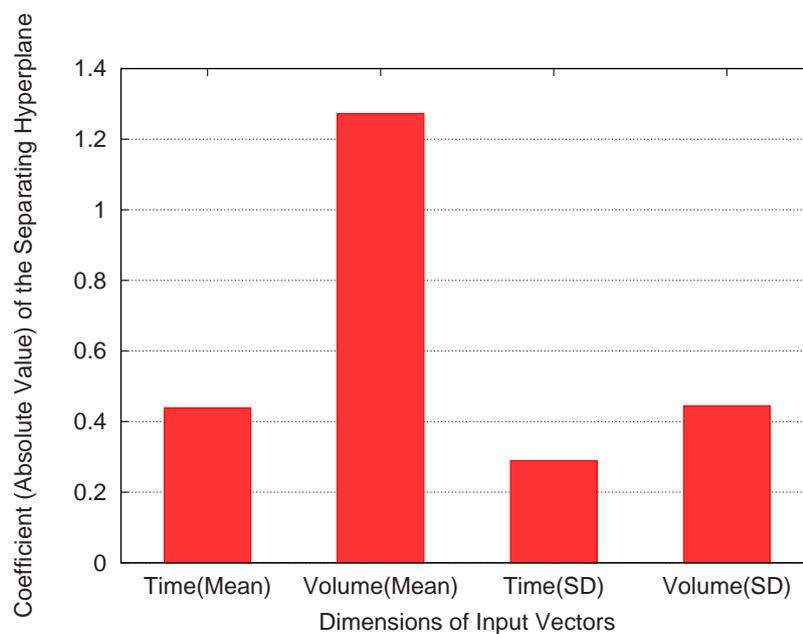


図 4.1: ハイハットシンバル 1 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

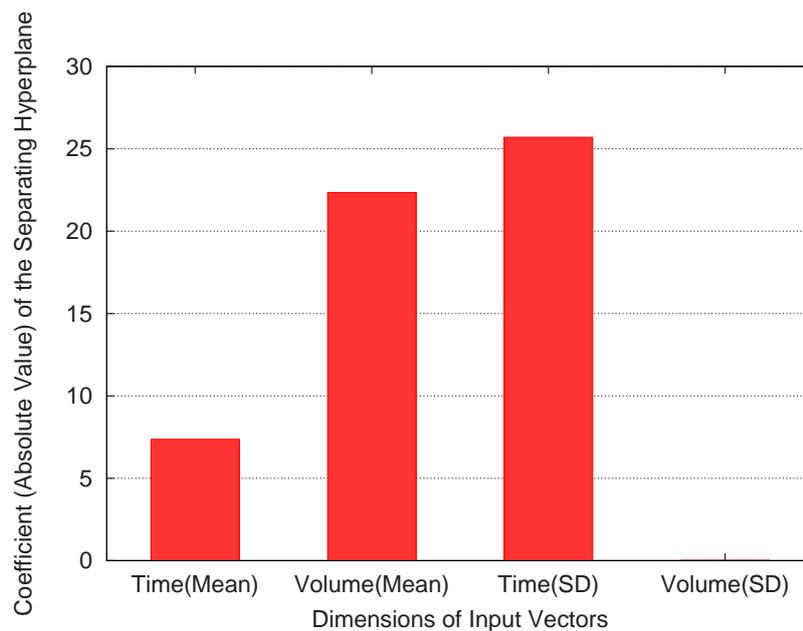


図 4.2: ハイハットシンバル 1.5 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

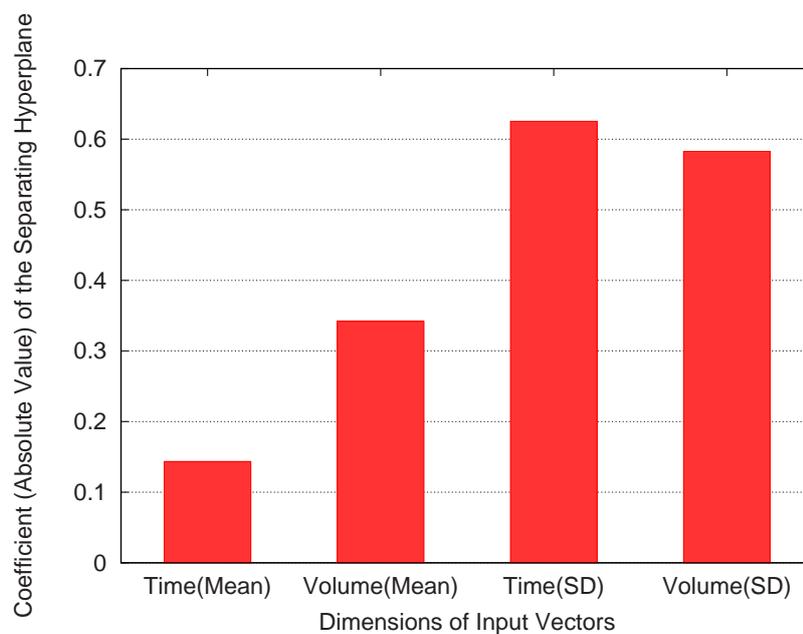


図 4.3: ハイハットシンバル 2 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

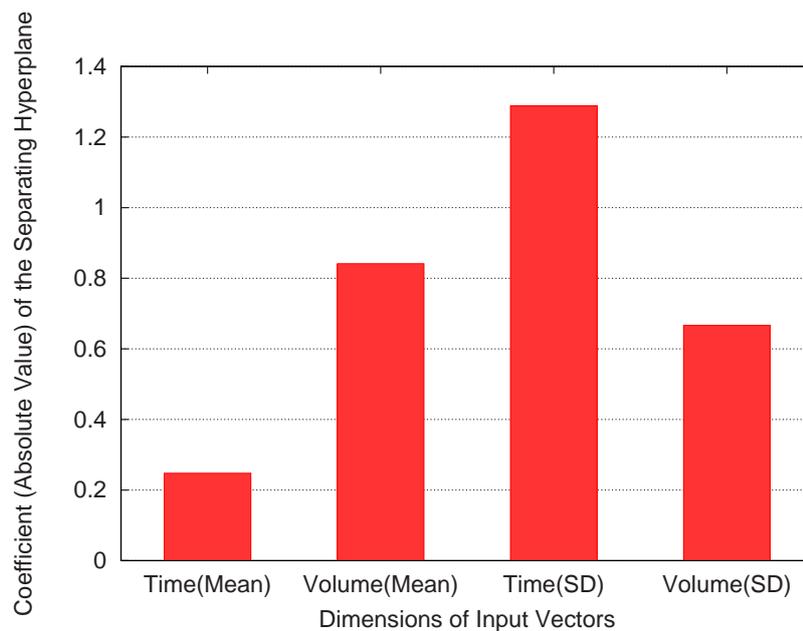


図 4.4: ハイハットシンバル 2.5 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

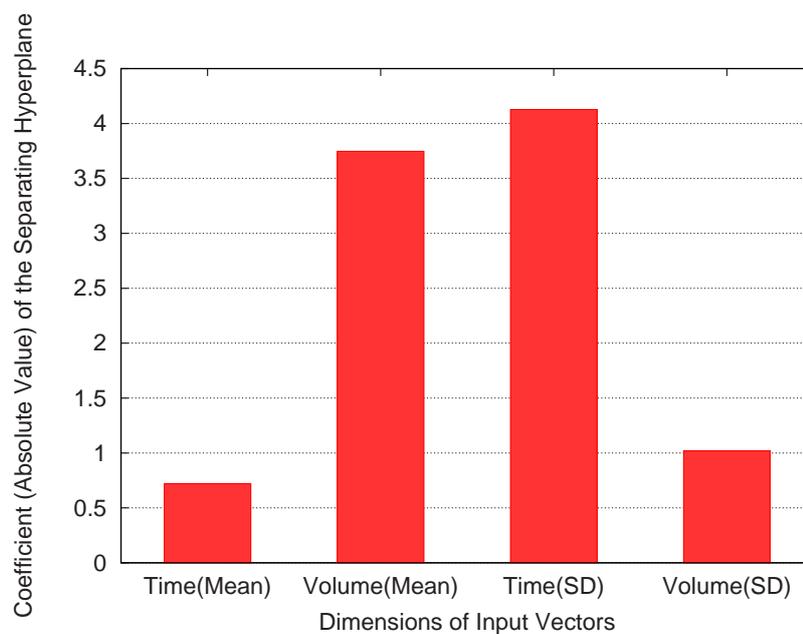


図 4.5: ハイハットシンバル 3 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

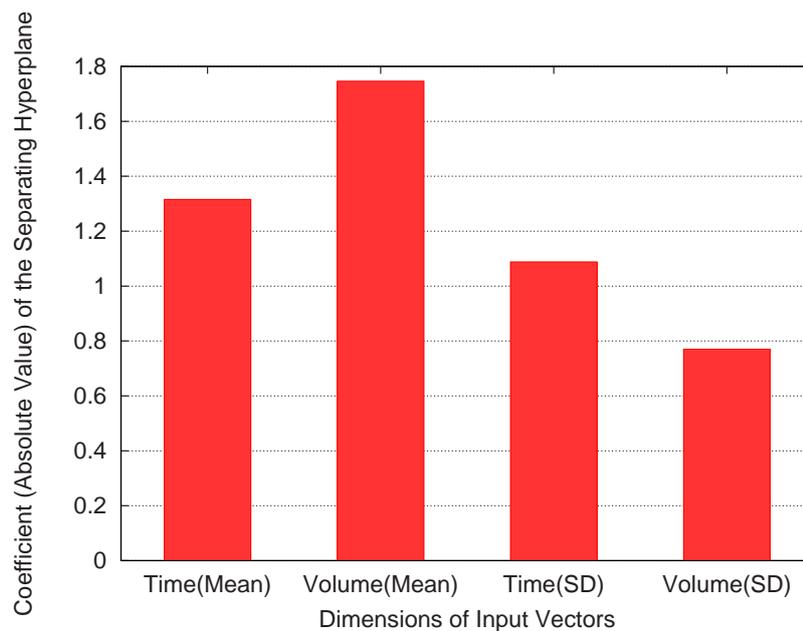


図 4.6: ハイハットシンバル 3.5 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

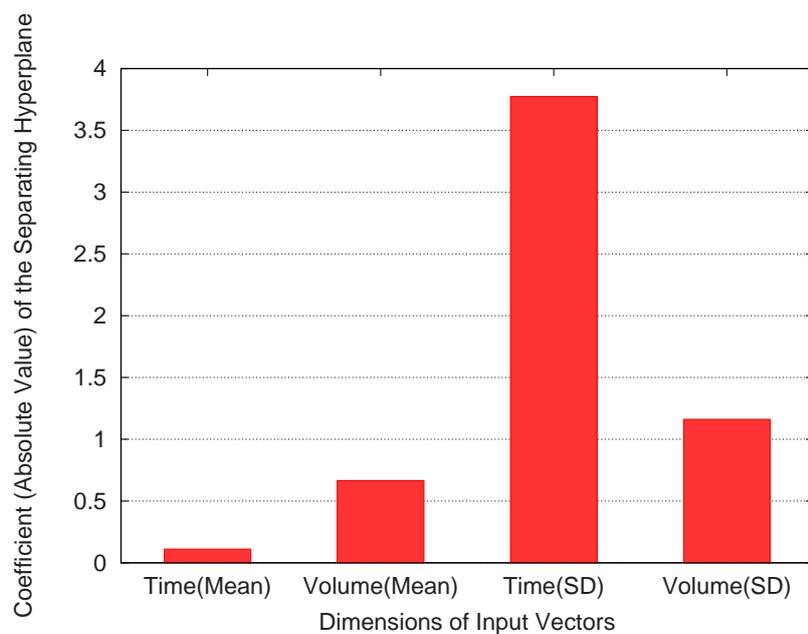


図 4.7: ハイハットシンバル 4 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

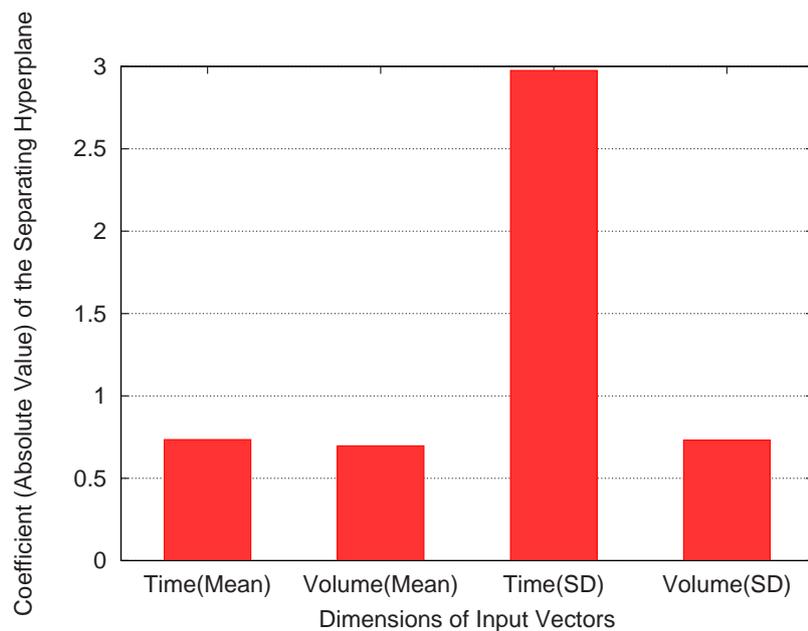


図 4.8: ハイハットシンバル 4.5 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

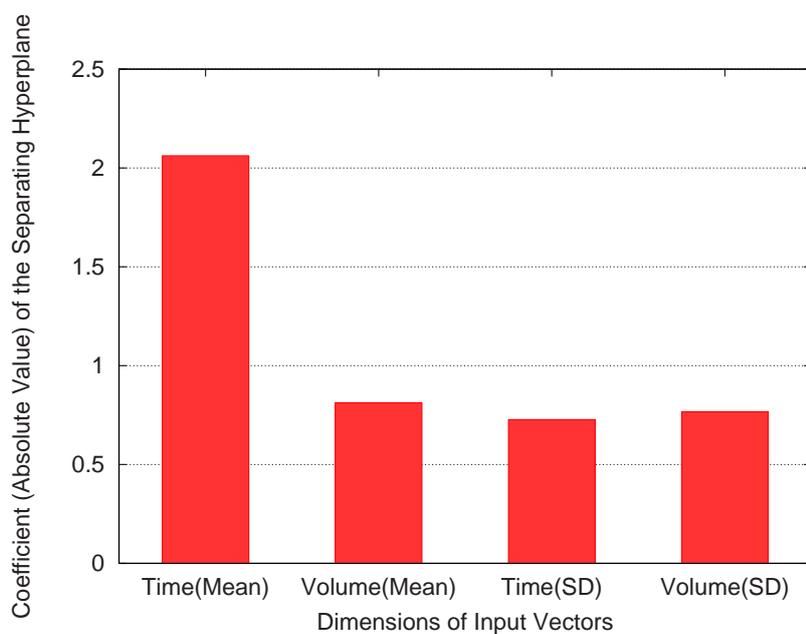


図 4.9: バスドラム 1 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

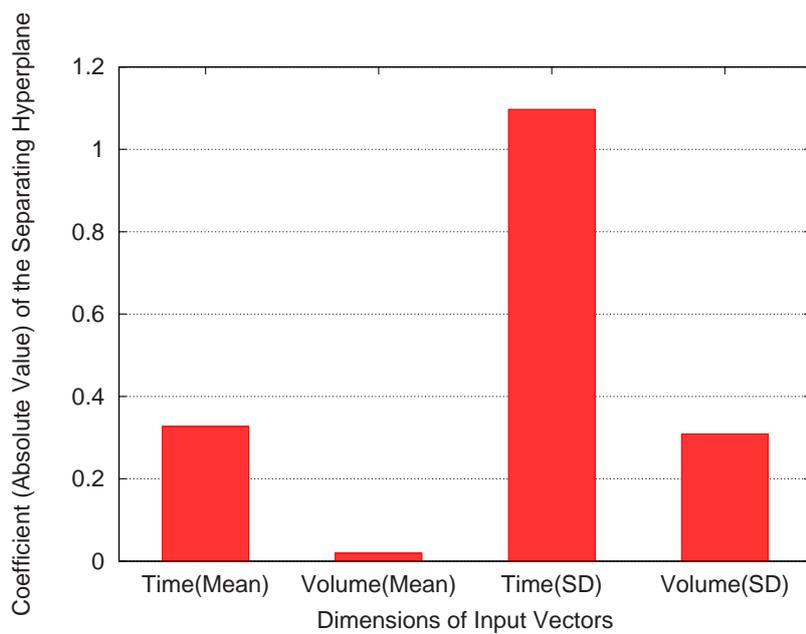


図 4.10: バスドラム 3 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

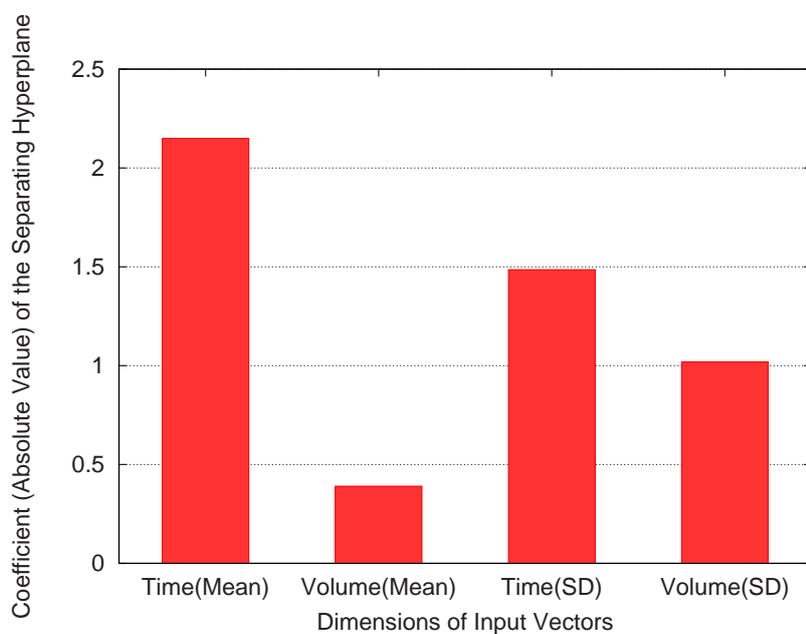


図 4.11: スネアドラム 2 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

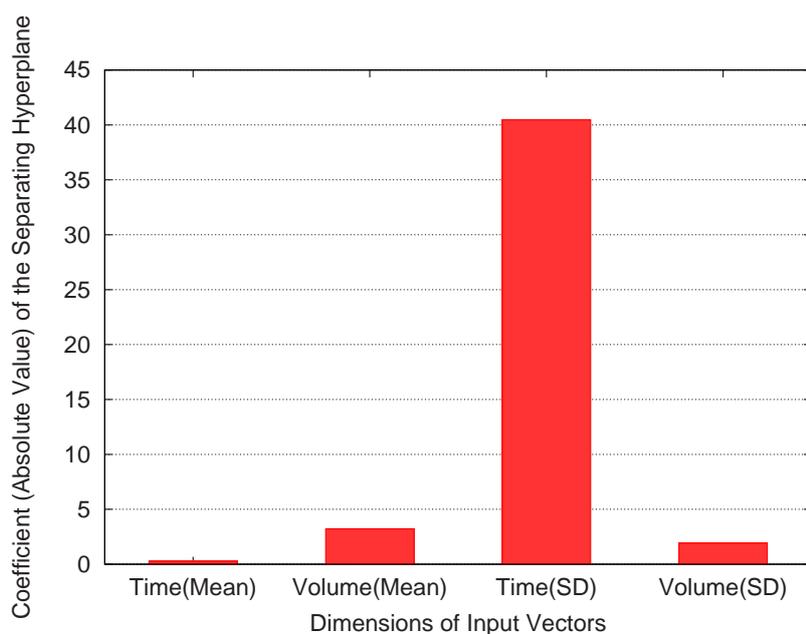


図 4.12: スネアドラム 4 拍目 分離超平面の方程式における各次元の係数の絶対値の比較

表 4.4: 分離超平面の方程式における各次元の係数の絶対値の順位獲得回数

絶対値の 順位	打点時刻ずれの 平均	打点音量の 平均	打点時刻ずれの 標準偏差	打点音量の 標準偏差
1 位	2	2	8	0
2 位	2	5	1	4
3 位	2	3	1	6
4 位	6	2	2	2

4.2.4 考察 — 重要な特徴の判別

ここで、図 4.1 ~ 図 4.12 をまとめて、各楽器・各打拍においてそれぞれの次元の絶対値が何番目に大きいものであったかの順位について、その順位の獲得回数を表 4.4 に示す。

表 4.4 に示した通り、分離超平面の係数の絶対値が最大（1 位）となった回数が「8 回」となり 4 種の次元の中で圧倒的に多かったのは打点時刻ずれの標準偏差であった。このことより、

- ルースとタイトというグルーブ感を分類する最も重要な特徴は打点時刻ずれの標準偏差である

ということが判明した。特にハイハットシンバル 4 拍目（図 4.7）・ハイハットシンバル 4.5 拍目（図 4.8）・バスドラム 3 拍目（図 4.10）・スネアドラム 4 拍目（図 4.12）などでは、打点時刻ずれの標準偏差が他の次元の係数の絶対値を大きく引き離す値をとっており、これらの楽器・打拍においてはグルーブ感による差が打点時刻ずれの標準偏差に特に顕著に現れていることを示している。2.2.2 節で述べた先行研究では、人間の目によるヒューリスティックな数値の比較により、「ルースの打点時刻の分散がタイトの打点時刻の分散よりも大きいものが多かった」という知見が得られていたが、今回の実験結果はその知見をシステムティックに裏付ける形となった。

次に、1 位の獲得回数では打点時刻の平均と打点音量の平均がそれぞれ「2 回」で並んでいるが、2 位の獲得回数では打点音量の平均が大きく上回っているため、打点時刻ずれの標準偏差の次にグルーブ感の分類で重要な役割を担っている特徴は打点音量の平均であると言える。

一方，打点時刻ずれの平均は，バスドラム 1 拍目（図 4.9）とスネアドラム 2 拍目（図 4.11）で係数の絶対値が 1 位となったものの，他の楽器・打拍ではあまり係数の絶対値が大きくなることは無かった．ただし，これは仮想メトロノームをハイハットシンバルの打点時刻を基準として算出していることも影響していると考えられるため，打点時刻ずれの平均がグルーヴ感の分類に寄与していないと一概に言い切ることはできない．

また，打点音量の標準偏差については，係数の絶対値が最も大きくなる楽器・打拍はひとつも無かった．特にハイハットシンバル 1.5 拍目（図 4.2）では他の次元の係数の絶対値に比べ遥かに小さい値をとってしまっているなど，グルーヴ感の分類にはあまり寄与していないことがわかった．

4.3 非線形 SVM による 4 次元特徴空間での分類と分類性能比較

以下では、ルース・タイトの両グルーブ感が含まれた 17 曲分のデータに対して、非線形 SVM による解析を行う。カーネルトリックを用いて非線形の分離超曲面を生成することにより、各楽器・各打拍におけるグルーブ感の分類を行う。そして、その分類結果を、4.2 節で行った線形 SVM による分離超平面による分類の際の結果と比較する。

4.3.1 実験手順

ここでの実験の手順は以下の通りである。

1. SVM に用いるデータの準備
2. 非線形 SVM に用いるカーネル関数の選定
3. クロスバリデーションによる非線形 SVM のパラメータチューニング
4. 非線形 SVM による分離超曲面の生成
5. 線形 SVM による分類結果との比較

また、SVM による分離超平面の生成は、4.2 節と同様に次の各楽器・各打拍ごとに行う。

- ハイハットシンバル
 - 1 拍目
 - 1.5 拍目
 - 2 拍目
 - 2.5 拍目
 - 3 拍目
 - 3.5 拍目
 - 4 拍目
 - 4.5 拍目

- バスドラム
 - 1 拍目
 - 3 拍目
- スネアドラム
 - 2 拍目
 - 4 拍目

なお，上記の「 拍目」という表記は，各小節内での拍の位置を示す

また，データを SVM に入力する際に設定するクラスラベル y_i についても 4.2 節と同様とし，グルーブ感によって以下のような値に設定する．

- ルースなグルーブ感の演奏から得たデータ： $y_i = -1$
- タイトなグルーブ感の演奏から得たデータ： $y_i = 1$

4.3.2 入力データの次元について

ここで解析を行うデータは，各曲ごとに 1 データずつ，計 17 個のデータとする．4.2 節と同様に，各データは，各曲における以下の 4 種の特徴の値を持つ 4 次元ベクトルとする．

- 打点時刻ずれの平均
- 打点音量の平均
- 打点時刻ずれの標準偏差
- 打点音量の標準偏差

4.3.3 カーネル関数の選定

ここでの実験では非線形 SVM による分類を行うため，2.3.3 節で述べたカーネルトリックの手法が必要となる．カーネルトリックを用いるためには，2.3.3 節で挙げたようないくつかのカーネル関数の中から使用するものを選定する必要がある．

ここでは、カーネル関数として以下の式で表される Radial Basis Function (RBF) を用いる。

$$K(\boldsymbol{x}, \boldsymbol{z}) = \exp(-\gamma \|\boldsymbol{x} - \boldsymbol{z}\|^2), \quad \gamma > 0 \quad (4.2)$$

カーネル関数を選定する際、最初に RBF を選ぶのは妥当な選択であるとされている [21]。数ある代表的なカーネル関数の中から今回 RBF を用いる理由は次の通りである。

- 設定すべきパラメータの数が少ない (γ のみ)。
- 線形関数や式 (2.28) のシグモイドカーネル関数による分類は、RBF カーネル関数による分類の特殊なケースであると見ることができる [22]。すなわち、RBF カーネル関数による分類は汎用性が高い。

なお、クロスバリデーションによる SVM パラメータの設定の際には、ソフトマージンアルゴリズムのパラメータ C に加え、RBF の式 (4.2) の中に現れる変数 γ についてもチューニングを行う必要がある。

4.3.4 実験結果

4.3.4.1 クロスバリデーションによるパラメータチューニングの結果

まず、4.2.3.1 節と同様に、過学習を避けるべくクロスバリデーションによる SVM のパラメータの設定を行った。ここでは、ソフトマージンアルゴリズムのパラメータ C および RBF の式 (4.2) の中に現れる変数 γ の 2 種類のパラメータについてチューニングを行った。それぞれのパラメータの値を、

- $C : 2^{-5}, 2^{-4}, \dots, 2^{15}$
- $\gamma : 2^{-15}, 2^{-14}, \dots, 2^3$

と等比数列的に変化させながら、各 C および γ の値を用いて 17 分割クロスバリデーションを逐一行い、クロスバリデーションの平均分類正解率が最も高くなった C と γ の値の組み合わせを採用した。

表 4.5: 各楽器・各打拍における非線形 SVM の最適パラメータ (4 次元特徴空間)

楽器	小節内の打拍位置	C	γ	クロスバリデーションの 平均正解率 (%)
ハイハットシンバル	1 拍目	128.0	0.001953125	76.4706
	1.5 拍目	64.0	0.0625	88.2353
	2 拍目	1.0	0.125	76.4706
	2.5 拍目	16.0	0.015625	88.2353
	3 拍目	16384.0	0.00048828125	100.0000
	3.5 拍目	2048.0	0.00048828125	94.1176
	4 拍目	8192.0	0.00048828125	70.5882
バスドラム	4.5 拍目	1.0	0.5	94.1176
	1 拍目	2048.0	0.00048828125	88.2353
スネアドラム	3 拍目	8.0	0.015625	88.2353
	2 拍目	1024.0	0.001953125	82.3529
	4 拍目	4096.0	0.03125	82.3529

クロスバリデーションを用いて設定された各楽器・各打拍における C および γ の値と、その際のクロスバリデーションの平均分類正解率を、表 4.5 に示す。以下の RBF カーネルによる非線形分類実験では、ここで設定された C および γ の値を最適パラメータとして用いるものとする。

4.3.4.2 分離超曲面の生成と分類正解率

表 4.5 の C および γ の値を用いた非線形 SVM によって、ルースとタイトのグルーブ感を分類する分離超曲面を、各楽器・各打拍ごとに生成した。

生成された分離超曲面によるグルーブ感の分類正解率を、表 4.6 に示す。表 4.6 を見ると、分類正解率が最も低かったのはハイハットシンバル 1 拍目・2 拍目の 82.3529% で、最も高かったのはハイハットシンバル 3 拍目・ハイハットシンバル 4.5 拍目・スネアドラム 4 拍目のいずれも 100.0000% であった。また、その他の楽器・打拍についても、軒並み 80% 台後半～90% 台の高い正解率となっている。

表 4.6: 非線形 SVM の分離超曲面による分類正解率

楽器	小節内の打拍位置	分離超曲面による 分離正解率 (%)
ハイハットシンバル	1 拍目	82.3529
	1.5 拍目	94.1176
	2 拍目	82.3529
	2.5 拍目	94.1176
	3 拍目	100.0000
	3.5 拍目	94.1176
	4 拍目	88.2353
バスクラム	1 拍目	88.2353
	3 拍目	88.2353
スネアドラム	2 拍目	82.3529
	4 拍目	100.0000

4.3.5 考察 — 線形分類と非線形分類の性能比較

ここで、4.2 節の表 4.2 にて示した線形 SVM の分離超平面による分類正解率と、前述の表 4.6 にて示した非線形 SVM の分離超曲面による分類正解率とをまとめて比較したものを、表 4.7 に示す。

表 4.7 を見ると、線形 SVM から非線形 SVM へと拡張したことにより、ハイハットシンバル 2 拍目とハイハットシンバル 4.5 拍目の分類正解率が上昇した一方で、ハイハットシンバル 1.5 拍目については逆に分類正解率が下降してしまったことがわかる。バスクラム・スネアドラムについては、分類正解率に変化は無かった。

全体として、これらの分類正解率の変化は微々たるものであり、今回の 4 次元の入力ベクトルによる分類においては、線形 SVM でも十分な分類がなされていたことがわかった。実際、4.2 節でも述べた通り、線形 SVM による分類であっても 80% 台後半～90% 台の高い正解率となっている。ルースとタイトの分類結果を応用する場合に簡潔なモデルの構築が必要となる場合は、線形 SVM によるシンプルな線形分類で十分な効果が得られるということを示している。

表 4.7: 線形 SVM と非線形 SVM の分類正解率の比較

楽器	小節内の打拍位置	線形 SVM の 分離超平面による 分離正解率 (%)	非線形 SVM の 分離超曲面による 分離正解率 (%)
ハイハットシンバル	1 拍目	82.3529	82.3529
	1.5 拍目	100.0000	94.1176
	2 拍目	76.4706	82.3529
	2.5 拍目	94.1176	94.1176
	3 拍目	100.0000	100.0000
	3.5 拍目	94.1176	94.1176
	4 拍目	88.2353	88.2353
	4.5 拍目	94.1176	100.0000
バスドラム	1 拍目	88.2353	88.2353
	3 拍目	88.2353	88.2353
スネアドラム	2 拍目	82.3529	82.3529
	4 拍目	100.0000	100.0000

4.4 非線形 SVM による 2 次元特徴空間での分類

4.2 節で行った実験の結果，特に打点時刻ずれの標準偏差と打点音量の平均がグルーブ感を分類する重要な要因となっていることが判明した．そこで以下では，この打点時刻ずれの標準偏差と打点音量の平均の 2 種類の特徴に注目し，ルース・タイトの両グルーブ感が含まれた 17 曲分のデータに対して非線形 SVM による詳細な解析を行う．カーネルトリックを用いて非線形分離曲線を生成し，これを 2 次元平面上に図示することにより，各楽器・各打拍におけるグルーブ感の分布の様子を調査する．

4.4.1 実験手順

ここでの実験の手順は以下の通りである．

1. SVM に用いるデータの準備
2. 非線形 SVM に用いるカーネル関数の選定
3. クロスバリデーションによる非線形 SVM のパラメータチューニング
4. 非線形 SVM による分離曲線の生成
5. 生成された分離曲線によるグルーブ感の分類の様子の調査

また，SVM による分離超平面の生成は，4.2 節・4.3 節と同様に次の各楽器・各打拍ごとに行う．

- ハイハットシンバル
 - 1 拍目
 - 1.5 拍目
 - 2 拍目
 - 2.5 拍目
 - 3 拍目
 - 3.5 拍目

- 4 拍目
- 4.5 拍目
- バスドラム
 - 1 拍目
 - 3 拍目
- スネアドラム
 - 2 拍目
 - 4 拍目

なお、上記の「 拍目」という表記は、各小節内での拍の位置を示す

また、データを SVM に入力する際に設定するクラスラベル y_i についても 4.2 節・4.3 節と同様とし、グルーブ感によって以下のような値に設定する。

- ルースなグルーブ感の演奏から得たデータ： $y_i = -1$
- タイトなグルーブ感の演奏から得たデータ： $y_i = 1$

4.4.2 入力データの次元について

ここで解析を行うデータは、各曲ごとに 1 データずつ、計 17 個のデータとする。4.2 節で行った実験の結果を踏まえ、各データは、各曲における以下の 2 種の特徴の値を持つ 2 次元ベクトルとする。

- 打点時刻ずれの標準偏差
- 打点音量の平均

4.4.3 カーネル関数の選定

ここでの実験では非線形 SVM による分類を行うため、2.3.3 節で述べたカーネルトリックの手法が必要となる。ここでは、4.3.3 節と同様、カーネル関数として以下の式で表される

Radial Basis Function (RBF) を用いる .

$$K(\mathbf{x}, \mathbf{z}) = \exp(-\gamma \|\mathbf{x} - \mathbf{z}\|^2), \quad \gamma > 0 \quad (4.3)$$

RBF 関数をカーネル関数として選定する理由は 4.3.3 節と同様である . クロスバリデーションによる SVM パラメータの設定の際には , ソフトマージンアルゴリズムのパラメータ C に加え , RBF の式 (4.3) の中に現れる変数 γ についてもチューニングを行う必要がある .

4.4.4 実験結果

4.4.4.1 クロスバリデーションによるパラメータチューニングの結果

まず , 4.2.3.1 節・4.3.4.1 節と同様に , 過学習を避けるべくクロスバリデーションによる SVM のパラメータの設定を行った . ここでは , ソフトマージンアルゴリズムのパラメータ C および RBF の式 (4.3) の中に現れる変数 γ の 2 種類のパラメータについてチューニングを行った . それぞれのパラメータの値を ,

- $C : 2^{-5}, 2^{-4}, \dots, 2^{15}$
- $\gamma : 2^{-15}, 2^{-14}, \dots, 2^3$

と等比数列的に変化させながら , 各 C および γ の値を用いて 17 分割クロスバリデーションを逐一行い , クロスバリデーションの平均分類正解率が最も高くなった C と γ の値の組み合わせを採用した .

クロスバリデーションを用いて設定された各楽器・各打拍における C および γ の値と , その際のクロスバリデーションの平均分類正解率を , 表 4.8 に示す . 以下の RBF カーネルによる非線形分類実験では , ここで設定された C および γ の値を最適パラメータとして用いるものとする .

4.4.4.2 分離曲線の生成と分類正解率

表 4.8 の C および γ の値を用いた RBF カーネルによる非線形 SVM によって , ルースとタイトのグルーブ感を分類する分離曲線を , 各楽器・各打拍ごとに生成した . 以下では , 非線形 SVM による分類性能に関する結果を示す .

表 4.8: 各楽器・各打拍における非線形 SVM の最適パラメータ (2 次元特徴空間)

楽器	小節内の打拍位置	C	γ	クロスバリデーションの 平均分類正解率 (%)
ハイハットシンバル	1 拍目	256.0	0.00048828125	82.3529
	1.5 拍目	4096.0	0.5	82.3529
	2 拍目	32.0	0.015625	76.4706
	2.5 拍目	1.0	4.0	70.5882
	3 拍目	32.0	0.015625	70.5882
	3.5 拍目	128.0	0.015625	76.4706
	4 拍目	8192.0	0.125	64.7059
バスドラム	4.5 拍目	8.0	1.0	100.0000
	1 拍目	2048.0	0.001953125	58.8235
スネアドラム	3 拍目	8.0	2.0	76.4706
	2 拍目	1024.0	0.5	52.9412
	4 拍目	256.0	0.0625	64.7059

生成した分離曲線によるグルーヴ感の分類正解率を、表 4.9 に示す。なお、具体的な分離曲線の様子は図 4.13 ~ 図 4.24 にて後に示す。

表 4.9 を見ると、分類正解率が最も低かったのはハイハットシンバル 3 拍目の 70.5882% で、最も高かったのはハイハットシンバル 1.5 拍目・ハイハットシンバル 4.5 拍目のいずれも 100.0000% であった。ハイハットシンバル 1.5 拍目については表 4.2 で示した線形 SVM の分離超平面による分類正解率でも 100.0000% となっており、グルーヴ感による演奏の差が顕著に現れている楽器・打拍であると言える。また、その他の楽器・打拍については 70% 台後半 ~ 80% 台後半の正解率となったものが多く、グルーヴ感によって演奏にある程度の差が生じているということは確認できた。

しかし、これらは表 4.2 で示した線形 SVM の分離超平面による分類正解率よりも低くなっているものが多い。つまり、線形から非線形に拡張したにもかかわらず次元数が削減されたことにより、結果として SVM の分類能力が低下していることがわかる。このことより、非線形 SVM による分類であっても 2 次元のみの特徴による分類はやや難しいということが示されたと言える。また、グルーヴ感をより正確に分類するためにはより多くの演奏データサ

表 4.9: 非線形 SVM の分離曲線による分類正解率

楽器	小節内の打拍位置	分離曲線による 分類正解率 (%)
ハイハットシンバル	1 拍目	82.3529
	1.5 拍目	100.0000
	2 拍目	76.4706
	2.5 拍目	88.2353
	3 拍目	70.5882
	3.5 拍目	76.4706
	4.5 拍目	100.0000
バスドラム	1 拍目	70.5882
	3 拍目	94.1176
スネアドラム	2 拍目	88.2353
	4 拍目	82.3529

サンプルを用いて、より高次元のベクトルによる特徴空間での解析を行う必要があることを示唆しているということもできる。

4.4.4.3 分離曲線とグルーブ感の分類状況

ここで、各楽器・各打拍における各曲の演奏データの分布および分離曲線の様子を表したグラフを、図 4.13～図 4.24 に示す。図 4.13～図 4.24 において、グラフの縦軸は各曲における打点音量の平均値、グラフの横軸は各曲における打点時刻ずれの標準偏差をそれぞれ表している。また、図 4.13～図 4.24 の中で用いている各種記号等の凡例は表 4.10 の通りである。

なお、図 4.13～図 4.24 において表示してある分離曲線の幅に偏りが生じているが、これは、分離曲線をグラフ化する際に以下の手順によってグラフを作成しているためである。

1. 演奏データの値に応じてグラフの表示範囲を決定する。
2. グラフの表示範囲を 1000×1000 のピクセルに分割する。
3. 各ピクセルについて、RBF カーネルを用いた場合の分離曲線を表す判別関数 $f(x)$ の

表 4.10: 分離曲線を表す図の凡例

意味	図中の表記
ルースなグルーブ感による演奏の曲データ	
タイトなグルーブ感による演奏の曲データ	
分離曲線 ($ f(\boldsymbol{x}) \leq 0.1$ として作図)	
ルースに分類される領域 ($f(\boldsymbol{x}) < -0.1$ として作図)	
タイトに分類される領域 ($f(\boldsymbol{x}) > 0.1$ として作図)	

方程式

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{x}_i \in SV^*} \alpha_i^* y_i \exp(-\gamma \|\boldsymbol{x}_i - \boldsymbol{x}\|^2) + b \quad (4.4)$$

の値を計算し，不等式 $|f(\boldsymbol{x})| \leq 0.1$ を満たすピクセルを緑色 () で着色して分離曲線の一部とする．

従って，分離曲線の表示上の幅は $f(\boldsymbol{x})$ の勾配の程度を表してはいるが，グルーブ感の分類にあたっては特に大きな意味を持つものではない．また，4.4.1 節で述べたクラスラベル y_i の設定により，判別関数 $f(\boldsymbol{x}) < 0$ となる領域では入力データはルースに分類され，判別関数 $f(\boldsymbol{x}) > 0$ となる領域では入力データはタイトに分類される．