

令和5年度 卒業論文

# ハウスミュージックの楽曲構成を決定する要因 とその法則性の分析

指導教員 北原鉄朗 教授

日本大学文理学部情報科学科

ウルフ ジャスティン智也

2024年2月 提出

# 概 要

ハウスミュージックはダンスミュージックジャンルの一つであり、ポピュラーミュージックとは異なる構成で作曲される。各パートが同じパターンを何度も繰り返す中、各パートが演奏を始めるタイミングを巧みに制御することで起承転結を形作る。このような楽曲構成について把握することは、楽曲を理解する上で重要である。しかし、ハウスミュージックの楽曲構成が具体的にどのような要因に起因して決定され、どのような法則性があるかは定かではない。本稿では、ハウスミュージックの楽曲構成を決定する要因及びハウスミュージックの楽曲構成における法則性を解明することを目的として実験を行う。

前述のハウスミュージックの楽曲構成方法に基づいて、ハウスミュージックの楽曲構成と音圧の関連性について予備実験を行った結果、音圧の変化タイミングとセクションの変化タイミングに相関関係があることを確認した。また、予備実験の考察から、音圧に加えて周波数とドラム音について楽曲構成との関連性を調べる実験を行った。

実験の結果、それぞれの関連性についていくつかの特徴が見られ、それらはセクション間での変化が無い要素と変化のある要素の2つに分けられた。例えばセクション間での変化が無い要素としては、周波数は異なるセクションの場合でも変化は少ない。また、ドラム要素もセクションが異なる場合でも変化は無かった。一方、セクション間での変化がある要素としては、音圧はセクションが異なる場合には値が異なった。このことから、ハウスミュージックではセクションごとの特徴を明確に区分しつつも、一定の特徴は維持することで楽曲としての統一感を

保っていると考えられる。

今後の展望として、楽曲データの拡充や、セクション定義の妥当性の検証、セグメンテーション情報や分析アルゴリズムなどの精度向上、他ジャンルの楽曲との比較分析などが挙げられた。

# 目 次

目 次	iii
図 目 次	vii
表 目 次	xi
第 1 章 序 論	1
1.1 背景 . . . . .	1
1.2 目的 . . . . .	1
1.3 構成 . . . . .	2
第 2 章 関連研究と本研究のアプローチ	3
2.1 関連研究 . . . . .	3
2.1.1 関連研究 1: Visualizing Music and Audio using Self-Similarity[1]	3
2.1.2 関連研究 2 : Music Structure Analysis Using a Probabilistic Fitness Measure and an Integrated Musicological Model[2]	4
2.1.3 関連研究 3 : All-In-One Metrical And Functional Structure Analysis With Neighborhood Attentions on Demixed Audio[3]	4
2.1.4 関連研究 4 : Deep Embeddings and Section Fusion Improve Music Segmentation[4]	5
2.1.5 関連研究 5 : Hierarchical Evaluation of Segment Boundary Detection[5]	5

2.1.6	関連研究 6 : Structural Segmentation of Musical Audio by Constrained Clustering[6]	6
2.1.7	関連研究 7 : Boundary Detection in Music Structure Analysis Using Convolutional Neural Networks[7]	6
2.1.8	関連研究 8 : Analyzing Song Structure with Spectral Clustering[8]	6
2.1.9	関連研究 9 : A Scape Plot Representation for Visualizing Repetitive Structures of Music Recordings[9]	7
2.1.10	関連研究 10 : Learning Multi-Level Representations for Hi- erarchical Music Structure Analysis[10]	7
2.1.11	関連研究 11 : Colorscore : MIDI を利用したクラシック楽曲 構造の可視化と圧縮表示 [11]	7
2.2	本研究のアプローチ	8
<b>第 3 章</b>	<b>予備実験</b>	<b>9</b>
3.1	予備実験の流れ	9
3.1.1	データセット	9
3.1.2	楽曲構成のセクションの定義	10
3.1.3	音源の前処理	10
3.1.4	各パート及び分離前の音源に対する音圧の計算	11
3.1.5	グラフの描画	11
3.2	結果	12
3.3	考察	14
3.4	今後の方針	15
<b>第 4 章</b>	<b>実験手法</b>	<b>17</b>
4.1	それぞれの実験の仮説・目的	17
4.1.1	周波数に関する分析	17

4.1.2	音圧に関する分析 . . . . .	17
4.1.3	ドラム音に関する分析 . . . . .	18
4.2	分析の流れ . . . . .	18
4.2.1	楽曲構成のセクションの定義 . . . . .	18
4.2.2	楽曲のセグメンテーション情報の取得と前処理 . . . . .	19
4.2.3	データセット . . . . .	20
4.2.4	周波数に関する分析 . . . . .	20
4.2.5	音圧に関する分析 . . . . .	22
4.2.6	ドラム音に関する分析 . . . . .	22
<b>第 5 章</b>	<b>分析結果</b>	<b>25</b>
5.1	結果・考察 . . . . .	25
5.1.1	周波数に関する分析 . . . . .	25
5.1.2	音圧に関する分析 . . . . .	36
5.1.3	ドラム音に関する分析 . . . . .	49
5.1.4	総合考察 . . . . .	51
<b>第 6 章</b>	<b>まとめ</b>	<b>57</b>
6.1	結論 . . . . .	57
6.2	今後の展望・課題 . . . . .	58
	<b>参考文献</b>	<b>61</b>



# 目 次

3.1	Selecao - Mark Knight, Shovell . . . . .	12
3.2	Together - Thomas Bangalter, DJ Falcon . . . . .	13
3.3	I Feel For You Star B Extended Remix - Bob Sinclar, Riva Starr, Mark Broom, Star B . . . . .	13
3.4	Phoenix - Daft Punk . . . . .	14
4.1	All-In-One でのセグメンテーション情報の例 . . . . .	19
5.1	セクション別の Spectral Centroid の平均値からなる箱ひげ図. 白丸 は四分位範囲の 1.5 倍を超えた外れ値を表す. 緑の三角は平均値を 表す. オレンジの線は中央値を表す. . . . .	27
5.2	セクション別の Spectral Centroid の平均値からなるバイオリングラフ	28
5.3	各パートにおけるセクション別の Spectral Centroid の平均値からな る箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. 緑 の三角は平均値を表す. オレンジの線は中央値を表す. . . . .	29
5.4	ベースにおけるセクション別の Spectral Centroid の平均値からなる 箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレ ンジの線は中央値を表す. . . . .	30
5.5	ドラムにおけるセクション別の Spectral Centroid の平均値からなる 箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレ ンジの線は中央値を表す. . . . .	31



5.6	その他のパートにおけるセクション別の Spectral Centroid の平均値 からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表 す. オレンジの線は中央値を表す. . . . .	32
5.7	ヴォーカルにおけるセクション別の Spectral Centroid の平均値か らなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレンジの線は中央値を表す. . . . .	33
5.8	各パートにおけるセクション別の Spectral Centroid の平均値からな るバイオリングラフ . . . . .	34
5.9	RMS が閾値を下回る区間を除外した後の各パートのセクション別 の有効区間 . . . . .	35
5.10	セクション別の RMS の平均値からなる箱ひげ図. 白丸は四分位範 囲の 1.5 倍を超えた外れ値を表す. 緑の三角は平均値を表す. オレ ンジの線は中央値を表す. . . . .	39
5.11	セクション別の RMS の平均値からなるバイオリングラフ . . . . .	40
5.12	各パートにおけるセクション別の RMS の平均値からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. 緑の三角は平均 値を表す. オレンジの線は中央値を表す. . . . .	41
5.13	ベースにおけるセクション別の RMS の平均値からなる箱ひげ図. 白 丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレンジの線は中 央値を表す. . . . .	42
5.14	ドラムにおけるセクション別の RMS の平均値からなる箱ひげ図. 白 丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレンジの線は中 央値を表す. . . . .	43
5.15	その他のパートにおけるセクション別の RMS の平均値からなる箱 ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレ ンジの線は中央値を表す. . . . .	44

5.16 ヴォーカルにおけるセクション別の RMS の平均値からなる箱ひげ 図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレンジの 線は中央値を表す. . . . .	45
5.17 各パートにおけるセクション別の RMS の平均値からなるバイオリ ングラフ . . . . .	46
5.18 セクション別に点を色分けして 3 パートの音圧バランスを 3 次元座 標空間にプロットしたもの . . . . .	47
5.19 図 5.18 の図についてセクション別に座標空間を分けたもの . . . . .	48
5.20 各セクションにおける各ドラム要素の発音回数からなる棒グラフ . .	51
5.21 キックの 8 小節ごとの発音回数を折れ線グラフとしたスパゲッティ プロット . . . . .	52
5.22 スネアの 8 小節ごとの発音回数を折れ線グラフとしたスパゲッティ プロット . . . . .	52
5.23 ハイハットの 8 小節ごとの発音回数を折れ線グラフとしたスパゲッ ティプロット . . . . .	53
5.24 ドラムパターンの変化点を分母とした場合の一致率のヒストグラム	53
5.25 セクションの変化点を分母とした場合の一致率のヒストグラム . . .	54
5.26 ドラムパターンの変化点を分母とした場合の一致が発生した楽曲に おける時間帯のヒストグラム . . . . .	54
5.27 セクションの変化点を分母とした場合の一致が発生した楽曲におけ る時間帯のヒストグラム . . . . .	55



# 表 目 次

3.1 予備実験に使用した楽曲 . . . . .	9
4.1 セグメンテーション情報のラベルの変換 . . . . .	20
4.2 実験に使用したハウスミュージックのアルバム . . . . .	21
5.1 セクション別のクラスカル・ウォリス検定の結果 . . . . .	26
5.2 セクション別のダンの多重比較検定の結果 (p 値) . . . . .	26
5.3 各パートのセクション別のクラスカル・ウォリス検定の結果 . . . . .	26
5.4 ベースのセクション別のダンの多重比較検定の結果 (p 値) . . . . .	36
5.5 ドラムのセクション別のダンの多重比較検定の結果 (p 値) . . . . .	36
5.6 その他のパートのセクション別のダンの多重比較検定の結果 (p 値) . . . . .	37
5.7 ヴォーカルのセクション別のダンの多重比較検定の結果 (p 値) . . . . .	37
5.8 セクション別のクラスカル・ウォリス検定結果 . . . . .	38
5.9 セクション別のダンの多重比較検定の結果 (p 値) . . . . .	38
5.10 各パートのセクション別のクラスカル・ウォリス検定結果 . . . . .	38
5.11 ベースのセクション別のダンの多重比較検定の結果 (p 値) . . . . .	49
5.12 ドラムのセクション別のダンの多重比較検定の結果 (p 値) . . . . .	49
5.13 その他のパートのセクション別のダンの多重比較検定の結果 (p 値) . . . . .	50
5.14 ヴォーカルのセクション別のダンの多重比較検定の結果 (p 値) . . . . .	50
5.15 ドラムパターンの変化点を分母とした場合の一致率の平均値 . . . . .	51
5.16 セクションの変化点を分母とした場合の一致率の平均値 . . . . .	55



# 第1章 序 論

本章では，本研究を行う背景と目的，本稿の構成について説明する．

## 1.1 背景

ハウスミュージックはダンスミュージックの一種であり，ポピュラーミュージックとは異なる構成で作曲される．ポピュラーミュージックでは，メロディーやコード進行といった音高に関する要素の変化によって楽曲に展開を与えるが，対してハウスミュージックでは，各パートが同じパターンを何度も繰り返す中，各パートが演奏を始めるタイミングを巧みに制御することで起承転結を形作る．このような楽曲構成について理解することは，楽曲の制作を行う場合やDJプレイを行う上で重要である．しかし，ハウスミュージックの楽曲構成を決定する要因や楽曲構成の法則性は定かではない．これらの要因や法則性について分析を行うことで，作曲やDJプレイを行う際の参考資料にすることが出来る．

## 1.2 目的

楽曲構成の分析手法の提案は多々存在する．例えば，SSM(Self-Similarity Matrix)を用いた楽曲構成の可視化 [1] や，確率と探索アルゴリズムを用いた楽曲構成推定 [2]，ニューラルネットワークによるオールインワン型楽曲分析システム [3] などがある．しかし，ハウスミュージックに特化した分析手法は提案されていない．

本稿では、ハウスミュージックの楽曲構成を決定する要因の調査、及びハウスミュージックの楽曲構成における法則性を解明することを目的として分析を行う。

### 1.3 構成

本稿の構成について述べる。まず、2章で関連研究と本研究のアプローチについて述べ、本研究の位置づけを明確にする。次に、3章で予備実験について述べる。次に、4章で実験内容について詳しく述べる。さらに、5章にて4章で説明した実験の結果を紹介し、それに対する考察を述べる。6章で本稿のまとめと今後の展望について述べて結論とする。

## 第2章 関連研究と本研究のアプローチ

本章では、本研究の関連研究と、それに対する本研究のアプローチについて説明する。

### 2.1 関連研究

#### 2.1.1 関連研究 1 : Visualizing Music and Audio using Self-Similarity[1]

この論文では、SSM(Self-Similarity Matrix)を用いて楽曲の音響的類似性を2Dで表現する視覚化手法を提案している。オーディオの各瞬間をパラメータ化し、特徴量を用いたベクトル間の類似性を計算し、それに基づいてグレースケールの視覚表現を構築することで、楽曲の構造やリズムの特徴を視覚的に捉えることが出来る。この方法は、音楽の構造分析やテンポ抽出に応用可能であり、異なる楽曲間の類似性や、単一曲内での音響的特徴を識別するのに有効であるとしている。この手法により、楽曲構成におけるそれぞれのコンポーネントの長さを把握することが出来るが、それぞれのコンポーネントがどのような役割を担っているのかまでは分からない。



### 2.1.2 関連研究2: Music Structure Analysis Using a Probabilistic Fitness Measure and an Integrated Musicological Model[2]

この論文では、音楽構造分析としてメル周波数ケプストラム係数、クロマ特徴、リズムグラムなどの音響特徴量を利用したSDM(Self-Distance Matrix)による音楽構造分析システムを提案している。これらの音響特徴量に基づいて探索アルゴリズムにより音楽の各セクションを識別し、音楽的知識モデルに基づいてラベル付けするシステムである。この手法は、既存の音楽分析手法の限界を克服し、関連研究1の課題であった、それぞれのコンポーネントの役割が不透明である問題を、音楽学的知識モデルを用いたラベリングで解決するアプローチを取っている。しかし、評価実験ではセクションの識別部分がシステムのボトルネックとなることが問題点として示されている。

### 2.1.3 関連研究3: All-In-One Metrical And Functional Structure Analysis With Neighborhood Attentions on Demixed Audio[3]

この論文では、トランスフォーマーモデルを用いたニューラルネットワークによるオールインワン型楽曲分析システムを提案している。このシステムでは、分離音源のスペクトログラムを入力データとして使用し、時間的な長期依存性を捉えるためにDilated Neighborhood Attentions、局所的な楽器依存性のためにNon-Dilated Attentionsを用いている。このシステムはあらゆるタスクにおいて他の類似モデルをパフォーマンスで上回っており、高い精度で楽曲構成を推定出来ることが示されている。

#### 2.1.4 関連研究4: Deep Embeddings and Section Fusion Improve Music Segmentation[4]

この論文では, Laplacian Structural Decomposition をベースとした楽曲セグメンテーション手法が提案されている. 従来の手法と異なる点として, 学習に使用されている音響特徴量をディープラーニングによる音響埋め込みで置き換えている. 例えば, 楽曲におけるリピートセクションを識別するために, CQT(定Q変換)特徴量に加えて音楽自動タギングモデルからのディープ埋め込みを付け加えることで, 特徴量を補強している. これにより, マルチレベルでの楽曲セグメンテーションの精度が大幅に向上したことが示されている.

#### 2.1.5 関連研究5: Hierarchical Evaluation of Segment Boundary Detection[5]

この論文では, 音楽のセグメントの境界検出における階層的評価手法が提案されている. この手法は従来の単層的なアノテーションに基づく手法とは異なり, 音楽の階層構造 (短いフレーズ, 節, サビや間奏などの大きなセクションといった包括関係にある要素からなる構造) を考慮に入れている. 具体的に, Tree Measures と呼ばれる木構造を用いたフレームワークを提案している. これは, 音楽の階層的な構造を考慮し, 異なるセグメント間の関連性や差異を数値化し, それをランク付けすることで, 音楽の境界点の正確性を測定している. フレームワイズな類似性に基づくランキング問題として境界検出を評価する新しいフレームワークである. これにより, 単層的なアノテーションだけでなく, 階層的なアノテーションの比較も可能になっている.

### 2.1.6 関連研究6: Structural Segmentation of Musical Audio by Constrained Clustering[6]

この論文では、制約付きクラスタリングによる楽曲の構造的セグメンテーションの手法が提案されている。音楽の構造を階層的に把握し、隠れマルコフモデルを用いて楽曲オーディオフレームを特定の状態にラベル付けし、これらをクラスタリングする手法である。特に、時間的な連続性を制約として組み込んだクラスタリングアルゴリズムが導入されており、これにより音楽のセグメント間の境界をより適切に識別出来るようになっている。

### 2.1.7 関連研究7: Boundary Detection in Music Structure Analysis Using Convolutional Neural Networks[7]

この論文では、音楽構造分析におけるセグメント境界検出に、畳み込みニューラルネットワークを用いたアプローチを提案している。メルスペクトログラムを入力として使用し、楽曲の異なるセクション間の境界を自動的に検出することを目的としている。この手法は既存の技術と比較して、境界検出のF値で優れた結果を示しており、より正確にセグメントの境界を識別出来る手法である。

### 2.1.8 関連研究8: Analyzing Song Structure with Spectral Clustering[8]

この論文では、スペクトルのクラスタリングを用いた楽曲構成の分析手法を提案している。この手法は、SSMから導出された時間系列特徴を基に楽曲内の繰り返しパターンを分析し、大規模な構造の構築要素を捉える。特に、スペクトルグラフ理論の技術を応用し、低次元で繰り返し構造をエンコーディングし、異なる粒度レベルでの構造的コンポーネント間の階層的関係を明らかにする。

### 2.1.9 関連研究 9 : A Scape Plot Representation for Visualizing Repetitive Structures of Music Recordings[9]

この論文では、楽曲における繰り返し構造を可視化するための「スケーププロット」という新しい可視化方法を紹介している。この方法では、各音声セグメントを中心と長さで識別し、各点に色を割り当てることで、セグメントの繰り返し度合いと異なるセグメント間の関係を視覚的に表現している。この研究は、楽曲の繰り返し構造を階層的かつ直感的に理解するための有効な手法を提供し、音楽録音の構造的特性に新たな洞察をもたらすことが期待されている。

### 2.1.10 関連研究 10 : Learning Multi-Level Representations for Hierarchical Music Structure Analysis[10]

この論文では、音楽構造分析のための階層的なディープラーニング表現を非教師あり学習する方法を提案している。特に、音楽の時間的及び階層的な構造を利用して、異なる時間スケールでの音楽セグメンテーションを可能にするディープ埋め込み表現を生成する。

### 2.1.11 関連研究 11 : Colorscore : MIDI を利用したクラシック楽曲構造の可視化と圧縮表示 [11]

この論文では、MIDI データを活用したクラシック楽曲の構造の可視化システムが提案されている。このシステムは特に大編成のオーケストラ楽曲の理解を支援することを目的としている。楽曲の各パートが担うメロディの役割とその変遷を分析し、これを基に音楽構造を視覚的に表現する。重要なメロディを強調し、縦方向と横方向の圧縮表示を用いて、楽曲の全体像を短時間で捉えることが出来る。他の関連研究と比較して、この研究は音楽構造の可視化において、メロディの役

割とその変化を直感的に理解出来る点で特に優れている。しかし、この手法はクラシック楽曲に特化しており、ハウスミュージックへの利用には適していない。

## 2.2 本研究のアプローチ

上記で紹介した関連研究のうち、関連研究 1～10 は音楽ジャンルを問わない可視化・分析手法であり、関連研究 11 は大編成オーケストラのクラシック演奏に向けた可視化手法である。これらの関連研究の中に、ハウスミュージックに特化した手法は存在しない。そこで、本研究では対象をハウスミュージックに限定し、ハウスミュージックに特化した分析を行う。

1.1 で述べたように、ハウスミュージックはポピュラーミュージックとは異なり、基本的にメロディーやコード進行など音高に関連する要素が楽曲を通じて変わらない特徴がある。その代わり、同じパターンを繰り返す中、各パートが演奏を始めるタイミングを巧みに制御することで楽曲を構成するという特徴がある。このように各パートの音が重なり合うことによって楽曲が形作られるため、音圧の変化が楽曲の構造に関連していると考えることが出来る。したがって、ハウスミュージックの楽曲構成と音圧の関連性について調べることで、楽曲構成を決定する他の要因や、その法則性について足掛かりが得られると考えた。そこで、予備実験として、ハウスミュージックにおける楽曲構成と各パートの音圧の関連性を調査する実験を行なった。

## 第3章 予備実験

本章では予備実験について説明する．予備実験では，ハウスミュージックの楽曲に対して各パートの音圧と楽曲構成の関連性について調査するため，比較的単純な方法でハウスミュージックの楽曲の分析実験を行う．音源分離技術を使ってドラム，ベース，ヴォーカル，その他の音源に分離し，各音源の音圧を可視化し，そこから読み取れる特徴について考察する．

### 3.1 予備実験の流れ

#### 3.1.1 データセット

実験に使用する楽曲として，著名なハウスミュージック楽曲を4曲選定した．選定した楽曲を表 3.1 に示す．

表 3.1: 予備実験に使用した楽曲

楽曲名	アーティスト名
Selecao	Mark Knight, Shovell
Together	Thomas Bangalter, DJ Falcon
I Feel For You Star B Extended Remix	Bob Sinclar, Riva Starr, Mark Broom, Star B
Phoenix	Daft Punk

### 3.1.2 楽曲構成のセクションの定義

ハウスミュージックを始めとするダンスミュージックにおける楽曲構成の呼称に明確な定義は存在しないが、本稿では以下の4つのセクションを定義する。

- イントロ：曲の導入部分
- ドロップ：曲における盛り上がり部分
- ブレイク：曲の盛り上がり部分以外の部分
- アウトロ：曲の終結部分

ドロップは、ポピュラーミュージックにおけるサビに相当し、ブレイクは、サビの後の間奏やサビの前のAメロ・Bメロに近い位置づけである。ハウスミュージックなどの音楽では、上でも述べたように、メロディやコード進行を変えるのではなく、新たなパートを加えたりして楽曲のメリハリを作ることが多い。そのため、ここで述べたような楽曲構成は、音圧に現れる可能性が高いといえる。

### 3.1.3 音源の前処理

近年の音源分離技術の発展は目覚ましく、ミックス済みの音源からドラム、ベース、ヴォーカルなどの各パート音源を高精度に取り出すことが出来る。そこで、各パートの音源を取り出し、パート別の音圧についても合わせて分析する。

そこで、各パートの音源を音源分離システムである Demucs[13] を用いて生成する。Demucs は、Meta の研究部門である Meta Research が開発を行っているオープンソースの音源分離モデルである。Demucs は、音源分離コンテスト「Music Demixing Challenge 2021」で優勝していることから、精度の高い音源分離が期待出来る。Demucs は、入力として任意の音源ファイルを渡すと、4つのパート(ドラム、ベース、ヴォーカル、その他)に音源を分離して出力する。本稿の実験では、

入力する音源の形式は mp3 とし、出力される音源の形式は mp3、ビットレートは 320kbps、サンプリング周波数は 44.1kHz とする.

実行の際にオプションを付与する事で、6つのパート（ドラム、ベース、ヴォーカル、ピアノ、ギター、その他）に分離することも可能だが、ピアノの分離の精度が必ずしも十分とは言えないことと、楽曲構成の分析においては4パートで十分と考えたことから、本稿の実験では4パートへの分離を行う.

### 3.1.4 各パート及び分離前の音源に対する音圧の計算

音圧は、対象とする音源の波形の二乗平均平方根 (RMS) として計算する. 具体的に、波形  $y(t)$  における音圧  $S_i(t)$  は下の式で表される.

$$S_i(t) = \sqrt{\frac{1}{T} \sum_{\tau=0}^T y(t + \tau)^2 dt} \quad (3.1)$$

上の式において、 $i$  はパートを表す ( $i \in \{\text{drums, bass, vocals, others, mixed}\}$ ). RMS を計算する際の窓幅  $T$  を 65000 サンプル、時間分解能 ( $S(t)$  の時間間隔) を 16250 サンプルとした.

RMS は、Python の音楽分析用モジュール Librosa[12] を利用して算出する. その後、未分離音源の音圧を基準として、最大値が 1、最小値が 0 となるように各パートの音圧の値を正規化する.

### 3.1.5 グラフの描画

各パート（ドラム、ベース、ヴォーカル、その他）の音源から求めた音圧の時間変化  $S_{\text{bass}}, S_{\text{bass}}, S_{\text{vocals}}, S_{\text{others}}$  に加え、未分離音源の音圧  $S_{\text{mixed}}$  に基づいて楽曲を4つのセクション（イントロ、ドロップ、ブレイク、アウトロ）に分類して色付けして描画する. 各セクションの判定基準は以下の通りとする.



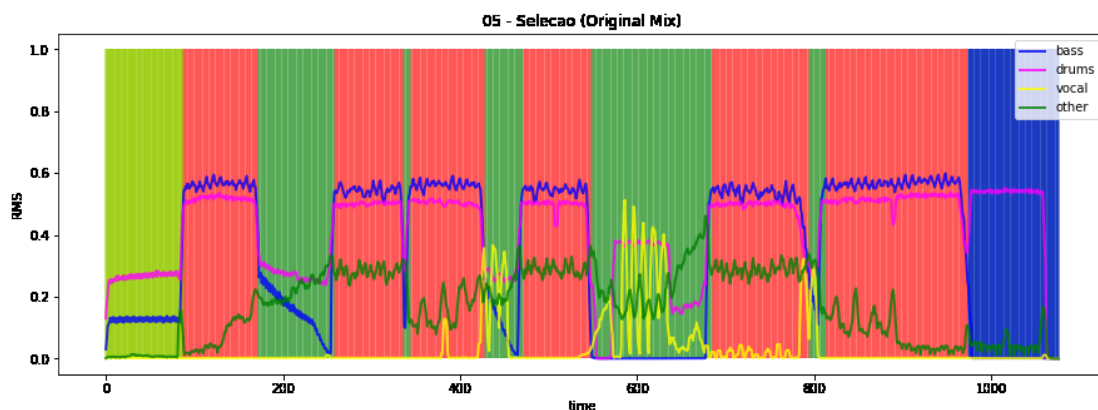


図 3.1: Selecao - Mark Knight, Shovell

- イントロ (黄) : 曲の開始時刻から, 最初に  $S_{\text{mixed}}$  が 0.85 を超える時刻までの区間
- ドロップ (赤) :  $S_{\text{mixed}}$  が 0.85 を超えている区間
- ブレイク (緑) :  $S_{\text{mixed}}$  が 0.85 を下回っている区間 (イントロ, アウトロに該当する区間を除く)
- アウトロ (青) :  $S_{\text{mixed}}$  が 0.85 を超える最後の時刻から, 曲の終了時刻までの区間

この基準は, サビに相当するドロップは, 基本的に他のセクションよりも音圧が大きいという仮説に基づく. 0.85 という閾値は, いくつかの楽曲を試しながら実験的に決定した.

## 3.2 結果

分析結果を図 3.1～図 3.4 に示す.

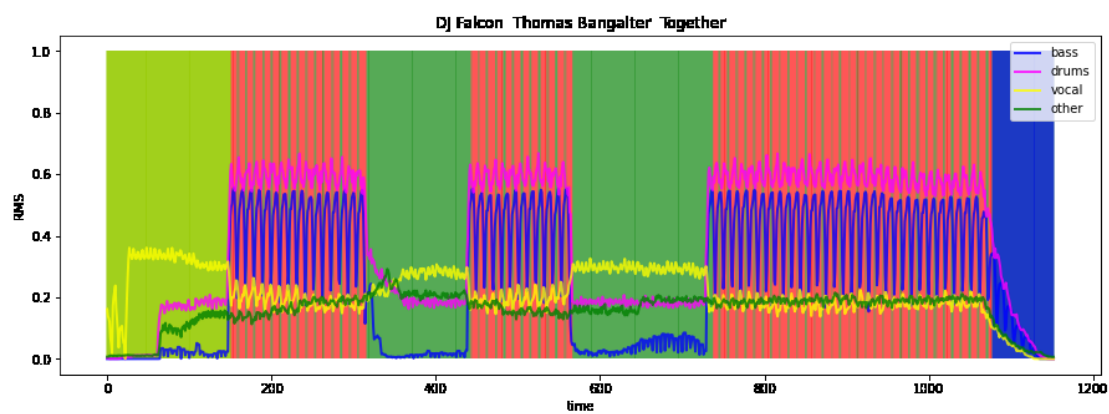


図 3.2: Together - Thomas Bangalter, DJ Falcon

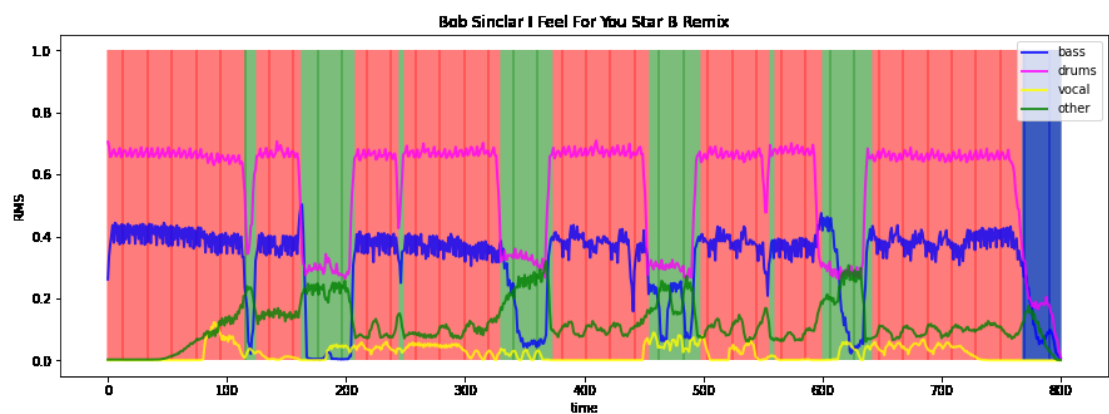


図 3.3: I Feel For You Star B Extended Remix - Bob Sinclar, Riva Starr, Mark Broom, Star B

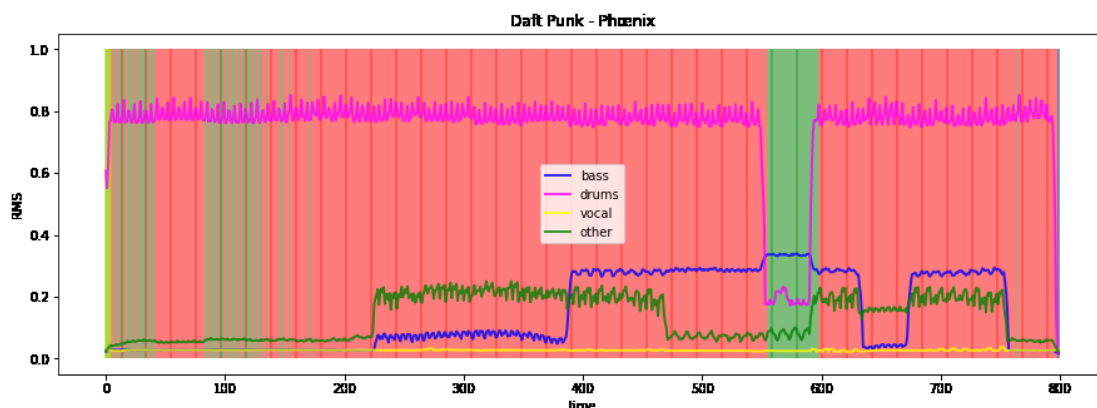


図 3.4: Phoenix - Daft Punk

### 3.3 考察

図 3.1 と図 3.2 の楽曲に関しては、4つのセクションが適切に分かれており、また、セクションが切り替わる場面では、各パートの音圧の増減がある事も読み取れる。例えば図 3.1 の楽曲では、イントロから最初のドロップに切り替わる際に、ドラムとベースの音圧が増加している。また、楽曲の中盤では、ドロップが終了したタイミングでドラムとベースが離脱し、代わりにヴォーカルが入り、次のドロップで再度ヴォーカルが離脱し、変わりにドラムとベースが入るといふ、音圧の入れ替わりがあることが確認出来る。

一方、図 3.3 の楽曲はイントロが、図 3.4 の楽曲はイントロとアウトロが消失している。特に図 3.4 の楽曲は、一部区間を除いて高音圧のドラムが鳴り続けており、ドラムの音圧だけで曲全体の音圧に大きな影響を及ぼしている。このことから、ハウスミュージックのセクションを判断する上で、特にイントロ、アウトロにおいては音圧のみでは考慮する要素が不十分であると言える。

また今回使用した4つの楽曲すべてにおいて、ドロップでは必ずドラムとベースが鳴っており、この2つのパートが音圧に大きく影響していると考えられる。また、ブレイクは大半でベースの音圧が低く、ドロップからブレイクへ切り替わる

タイミングで低音が抜ける傾向があることが分かる。

また、ドラムとベースがセクションの切り替わりに関与していることが多いのに対し、ヴォーカル及びその他のパートはセクションの切り替わりに関与していることが少ない。このことから、ハウスミュージックの楽曲構成においては、ドラムとベースが楽曲構成における展開を左右する役割を果たしている可能性が高いと考えられる。

### 3.4 今後の方針

本予備実験では、各パートの音圧の描画及び未分離音源の音圧に基づくセクション（イントロ、ドロップ、ブレイク、アウトロ）の分類・色付けによるハウスミュージックの楽曲構成と音圧の関連性を調べた。

実験結果から、ハウスミュージックの楽曲における各パートの音圧が楽曲構成の展開に与える影響が分かった。例えば、ドロップに入る際はドラムとベースの音圧が増加する傾向があった。反対に、ドラムとベース以外のパートは楽曲構成の展開にそこまで関与しておらず、ハウスミュージックにおいては主にドラムとベースが楽曲構成の展開における役割を果たしていると考えられることが分かった。

また、予備実験を通じて現状のセクション推定の問題点が浮かび上がった。例えば、ドロップとブレイクの自動推定は音圧を参照する方法が効果的であったのに対して、イントロとアウトロは音圧が高い場合と低い場合が存在するため、音圧に基づく判定のみでは不十分であった。このことから、ハウスミュージックにおけるイントロ・アウトロとドロップの違いは音圧以外の要素を考える必要がある。

これらの結果から、以降の実験においてハウスミュージックの楽曲構成との関連性について調査すべき要素として、次の2点が挙げられる。まず、低周波数を鳴らすドラムとベースが楽曲構成の展開に関与していることから、ハウスミュージックの楽曲構成には周波数が関連していると考えられる。そこで、周波数とハ

ウスミュージックの楽曲構成の関連性について調べる実験を行う。また、ドラムは一般的に多数の要素(キック, スネア, ハイハットなど)からなるパートなので、ドラムとハウスミュージックの楽曲構成の関連性について更に調べる余地がある。そこで、ドラムについてもハウスミュージックの楽曲構成との関連性について調べる実験を行う。さらに、音圧に関しては、今回の予備実験では比較的簡単な方法で各パートの音圧とハウスミュージックの楽曲構成について分析を行ったが、実験に使用した楽曲の数が少ないことと、音圧に基づいて楽曲構成が展開されるという前提のもとでセクションを定義した点に問題がある。これらの問題点を解消する為に、楽曲データを増やした上で音圧とハウスミュージックの楽曲構成の関連性について調べる実験を行う。

## 第4章 実験手法

本章では、実験の手法について説明する。予備実験を受けて、周波数、ドラム要素、音圧の3点とハウスミュージックの楽曲構成の関連性について調査する実験を行う。

### 4.1 それぞれの実験の仮説・目的

#### 4.1.1 周波数に関する分析

予備実験では、ドロップにてドラムとベースの音圧が高まる傾向があった。また、ドラムとベースは低周波数を鳴らすパートであることから、ドロップでは低周波数が強まるという仮説を考えることが出来る。反対に、盛り上がり部であるドロップで低周波数が強まる傾向があれば、盛り上がり部でないブレイクでは低周波数が弱まる傾向があると考えることが出来る。また、予備実験から、楽曲を通じて各パートの音圧は変化することが分かったが、同様に周波数に関しても各パートごとに楽曲を通じて変化があるかについて確認する。特に、予備実験から、ドラムとベースはドロップで音圧が高まり、それ以外のセクションで音圧が下がることがあったが、周波数に関しても同様の変化があるか調べる。

#### 4.1.2 音圧に関する分析

予備実験を行った楽曲では、ドロップにおいてドラムとベースの音圧が増加する傾向が見られた。一方で、ブレイクではこれらの音圧が低下する傾向にあった。

尚、ヴォーカル及びその他のパートの音圧変化は、セクションの変化と相関性が無かった。しかし、これらの結果は予備実験に用いた4曲に限定されるため、楽曲数を増やして実験を行うことで、信頼性の高い結果を得ることを目指す。また、予備実験において、イントロとアウトロでの音圧が増加する場合も観測されたが、これは全ての楽曲に当てはまるわけでは無かった。従って、イントロとアウトロにおける音圧の傾向を確認する。

### 4.1.3 ドラム音に関する分析

予備実験においては、ドラムの音圧の変化とセクションの変化が同時に発生する傾向が観測された。この音圧の変化は、ドラムパターンの変化、すなわちドラム要素の変更に起因すると考えられる。このことから、ドラムパターンの変化とセクションの変化が相関関係にあるかどうかを検証する。加えて、相関関係が確認された場合、セクションごとにその相関関係に差異があるかどうかを調査する。

さらに、各ドラム要素の発音回数がセクションによって異なるかどうかを分析する。特に、ハウスミュージックに特有の四つ打ちリズムは、ドラム要素の中でもキックを指す。これを含めた各ドラム要素と楽曲構成の関連性についても分析する。

## 4.2 分析の流れ

### 4.2.1 楽曲構成のセクションの定義

ハウスミュージックにおける楽曲構成の呼称に明確な定義は存在しないが、本稿では予備実験と同様に以下の4つを定義する。

- イントロ：曲の導入部分

- ドロップ：曲における盛り上がり部分
- ブレイク：曲の盛り上がり部分以外の部分
- アウトロ：曲の終結部分

### 4.2.2 楽曲のセグメンテーション情報の取得と前処理

実験に用いる楽曲構成の情報として，All-In-One[3]を用いて楽曲のセグメンテーション情報を取得する．All-In-Oneの概要については2.1.3で説明した通りである．

図4.1にAll-In-Oneで得られるセグメンテーション情報の例を示す．

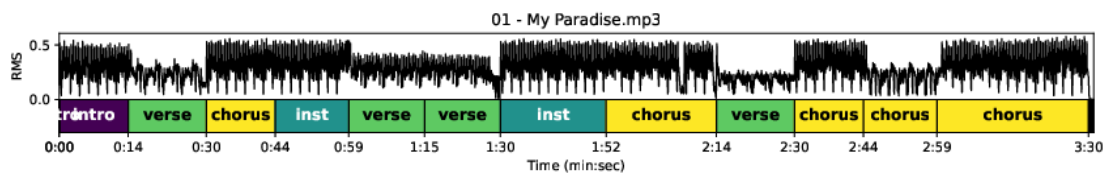


図 4.1: All-In-One でのセグメンテーション情報の例

All-In-Oneで取得出来るセグメンテーション情報では10種類のセグメンテーションのラベルが提供されているが，今回の実験では4.2.1の通り4つのセクションを定義しているので，定義した4種類のセクションのうち楽曲構成における役割が最も類似しているセクションに変換する．変換のルールを表4.1に示す．

また，All-In-Oneで得たセグメンテーション情報のうち，イントロとアウトロは概ね信頼性のある推定結果が得られたが，ドロップとブレイクは誤りと思われる結果が多かったため，

$$\text{Intro} \rightarrow [\text{Break}] \rightarrow (\text{Drop} \rightarrow \text{Break})^* \rightarrow [\text{Drop}] \rightarrow \text{Outro}$$

となるように手動で情報を補正する．例えば，同じ音が続いているのにも関わらずセクションが変化している場合は，どちらか一方の適切なセクションに統一する．ここで， $[]$ は省略可能， $()^*$ は0回以上の繰り返しを表す．



表 4.1: セグメンテーション情報のラベルの変換

元のセクション	変換後のセクション
start	intro
intro	intro
break	break
bridge	break
inst	break
verse	break
solo	drop
chorus	drop
outro	outro
end	outro

### 4.2.3 データセット

実験に使用する楽曲として、著名かつ歴史のあるハウスミュージックレーベルからリリースされているアルバムを7作品選定し、合計194曲を用意した。なるべく多種多様な作曲者の楽曲を収集するために、選択したアルバムはいずれもコンピレーションアルバムに限定した。また、アルバムの中に一部ハウスミュージックではない楽曲が存在したため、そのような楽曲は除外した。選択したアルバムを表4.2に示す。

### 4.2.4 周波数に関する分析

まず、楽曲データに対してAll-In-Oneを使用し、楽曲のセグメンテーション情報を取得する。次に、Demucsを使用して楽曲を4つのパート（ドラム、ベース、ヴォーカル、その他）に音源分離する。その後、これらのパートごとに周波数特徴量

表 4.2: 実験に使用したハウスミュージックのアルバム

レーベル名	アルバム名
De La Groove	3 Years On
Toolroom Records	Best Of Toolroom2018
Defected Records	Defected Presents Most Rated Ibiza 2022
Glitterbox	For Your Disco Pleasure Simon Dunmore
Glitterbox	Hotter Than Fire Molve Baptiste
Nervous Records	Nervous Records 30 Years
theBasement Discos	BEST OF 2022

を計算する．この際，周波数特徴量として Spectral Centroid を採用する．Spectral Centroid は，その周波数を分岐点に上下でエネルギーが均衡する点を示す特徴量であり，時間軸上の周波数バランスを効率的に分析出来る．各パートの Spectral Centroid は Librosa を用いて計算する．尚，無音区間での Spectral Centroid の計算を避けるため，RMS が特定の閾値を下回る区間は計算から除外する．

最後に，All-In-One で取得したセグメンテーション情報を用いて，各楽曲におけるセクション別の Spectral Centroid の平均値，及び各パートにおけるセクション別の Spectral Centroid の平均値を全楽曲データに対して算出し，その結果を箱ひげ図，バイオリン図で可視化する．

また，異なるセクション間のデータに有意差があるか検証する為に検定を行う．検定を行うにあたり，データの正規性についてダゴスティーノの  $K^2$  検定を用いて検証した結果，データは正規分布に従っていなかったため，非パラメトリック検定であるクラスカル・ウォリス検定を使用する．また，検定の結果， $p$  値が閾値 (0.05) を下回った場合，ダン検定を用いて事後検定を行い，どのセクションの組み合わせに有意差があるか調査する．

### 4.2.5 音圧に関する分析

予備実験と同様に、音圧を示す指標として波形の二乗平均平方根（RMS）を使用する．音圧に関して2種類の可視化を実施する．

一つ目の可視化では、4.2.4で用いた手法において、Spectral Centroid の代わりにRMSを使用し、その可視化を行う．

二つ目の可視化手順は以下の通りである．最初に、楽曲データに対して All-In-One を用いて楽曲のセグメンテーション情報を取得する．次に、Demucs を用いて楽曲を3パート（ドラム、ベース、ヴォーカル+その他）に音源分離する．その後、各パートに対して Librosa を用いてRMSを計算する．続いて、All-In-One で取得したセグメンテーション情報を用い、各パートのセクション別のRMSの平均値を全楽曲データに対して算出する．最後に、各楽曲のセクション別に3パートのRMSの平均値を3次元空間にプロットする．

また、4.2.4と同様の手法を用いて検定を行う．

### 4.2.6 ドラム音に関する分析

ドラムに関して2種類の可視化を実施する．

一つ目の可視化の手順を説明する．最初に、楽曲データに対して All-In-One を用いて楽曲のセグメンテーション情報を取得する．次に、Omnizart[14]を使用し、Demucsにより得たドラム音源をドラムMIDIファイルに変換し、その後PythonライブラリであるMidoを使用してドラムイベント情報を取得する．このドラムイベント情報を用いて3つのパターンで可視化を行う．一パターン目では、All-In-One で取得したセグメンテーション情報を用い、セクション別の各ドラム要素（キック、スネア、ハイハット）の個数を集計し、棒グラフで可視化する．二パターン目では、各ドラム要素の発音回数の蓄積を折線グラフによるスパゲッティプロットで可視化する．三パターン目では、All-In-One のセグメンテーション情報に含ま

れるテンポ情報を基に、楽曲内の8小節の長さを計算し、この長さを基準にして8小節ごとの各ドラム要素の発音回数を折線グラフによるスパゲッティプロットで可視化する。

二つ目の可視化手順は以下の通りである。一つ目の可視化と同様に、楽曲データに対して All-In-One を用いて楽曲のセグメンテーション情報を取得する。次に、Omnizart を使用し、Demucs により得たドラム音源をドラム MIDI ファイルに変換し、Mido を使用してドラムイベント情報を取得する。その後、ドラムパターンの変化点を検出する。この変化点は、全ドラムイベントの時間軸情報から各イベント間の間隔を取得し、その間隔の標準偏差を計算することで特定する。ドラムイベントがこの標準偏差を超える間隔で発生した場合、それをドラムパターンの変化と見なす。そのドラムパターンの変化点と All-In-One で取得したセクションの変化点に関して、変化点を整数に丸め、前後1秒の誤差を許容した一致率を計算する。この一致率を分布図で可視化する。また、変化点が一致した時間軸を基に、楽曲内での一致が発生した時間帯を楽曲の時間軸の割合で表示し、その分布を分布図で可視化する。



## 第5章 分析結果

本章では、実験の結果とその考察について述べる。

### 5.1 結果・考察

#### 5.1.1 周波数に関する分析

分析結果を図 5.1～図 5.9, 表 5.1～表 5.7 に示す。

図 5.1, 図 5.2 が示している通り、いずれのセクションも中央値と平均値は 4000Hz から 4500Hz 程度である。表 5.1 から、イントロとアウトロの組み合わせ以外はいずれもデータに有意差があるが、各セクションの平均値には最大で 500Hz 程度の差しかないことから、セクション別の周波数には、大きな差は無いと言える。4.1.3 で提起した仮説では、ドロップでは低周波数が強調され、ブレイクでは低周波数が弱まると考えたが、実際にはドロップとブレイクの周波数に顕著な差は見られなかった。また、ドロップとブレイクは比較的値の分散が小さく、Spectral Centroid が 4000Hz 程度にまとまりやすい傾向にある。対して、イントロとアウトロは値の分散が大きい傾向にある。

一方、パート別の可視化である図 5.3, 図 5.8 を見ると、セクションが異なる場合でも各パートの Spectral Centroid に大きさ違いは見られない。このことは、表 5.4～表 5.7 の検定結果が裏付けている。このことから、楽曲全体を通じて各パートが鳴らす周波数帯は一定を保ち、セクションの変化が各パートの周波数に大きな影響を与えていないことが分かる。

また、ベースはどのセクションでも Spectral Centroid が最も低く、対してドラムはどのセクションでも Spectral Centroid が最も高い。これは、キック以外のスネアやハイハットといった高周波数音の要素が影響していると考えられる。

しかし、図 5.3, 図 5.8 において、各パートの有効時間が異なる点を留意する必要がある。図 5.9 はセクションごとの各パートの有効時間の長さを可視化している。

表 5.1: セクション別のクラスカル・ウォリス検定の結果

統計量	70.71760398337165
p 値	$2.996 \times 10^{-15}$

表 5.2: セクション別のダンの多重比較検定の結果 (p 値)

	イントロ	ドロップ	ブレイク	アウトロ
イントロ	1.000	0.002914	$3.379 \times 10^{-9}$	0.986798
ドロップ	0.002914	1.000	0.033925	$5.755 \times 10^{-6}$
ブレイク	$3.379 \times 10^{-9}$	0.033925	1.000	$1.707 \times 10^{-13}$
アウトロ	0.986798	$5.755 \times 10^{-6}$	$1.707 \times 10^{-13}$	1.000

表 5.3: 各パートのセクション別のクラスカル・ウォリス検定の結果

パート	統計量	p 値
ベース	21.944235974164485	$6.700 \times 10^{-5}$
ドラム	15.194958854800916	0.001657
その他	12.616422142686588	0.005544
ボーカル	12.565717138080117	0.005676

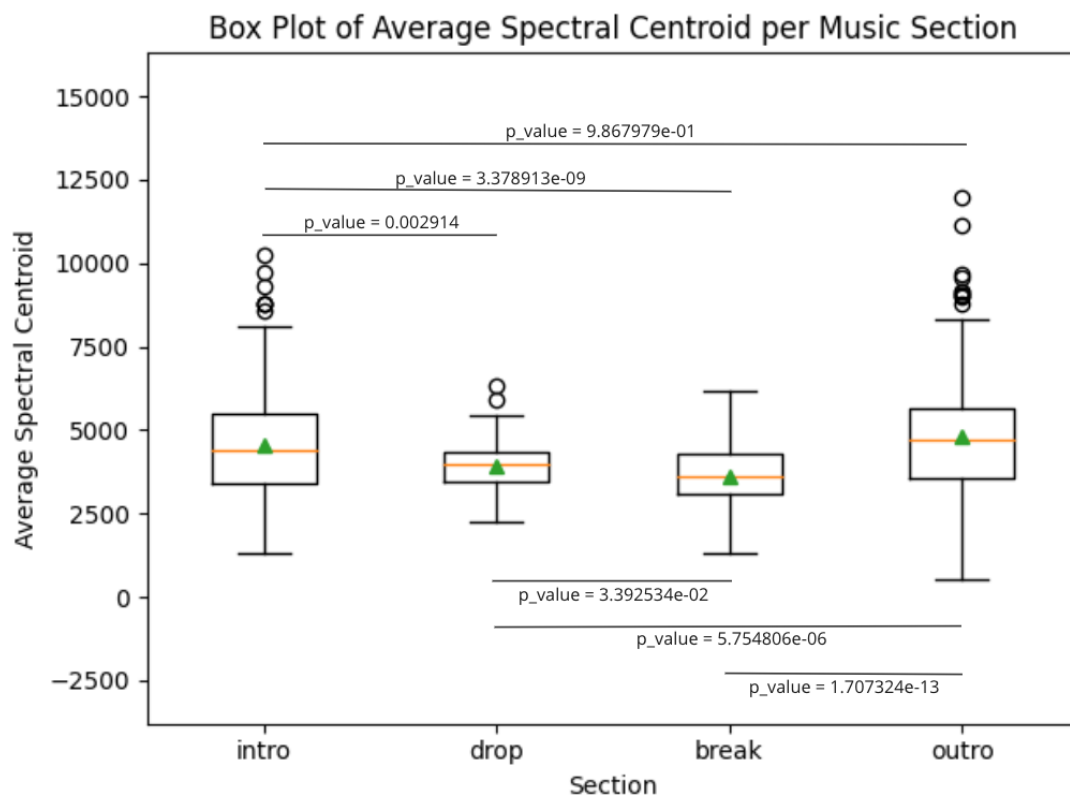


図 5.1: セクション別の Spectral Centroid の平均値からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. 緑の三角は平均値を表す. オレンジの線は中央値を表す.



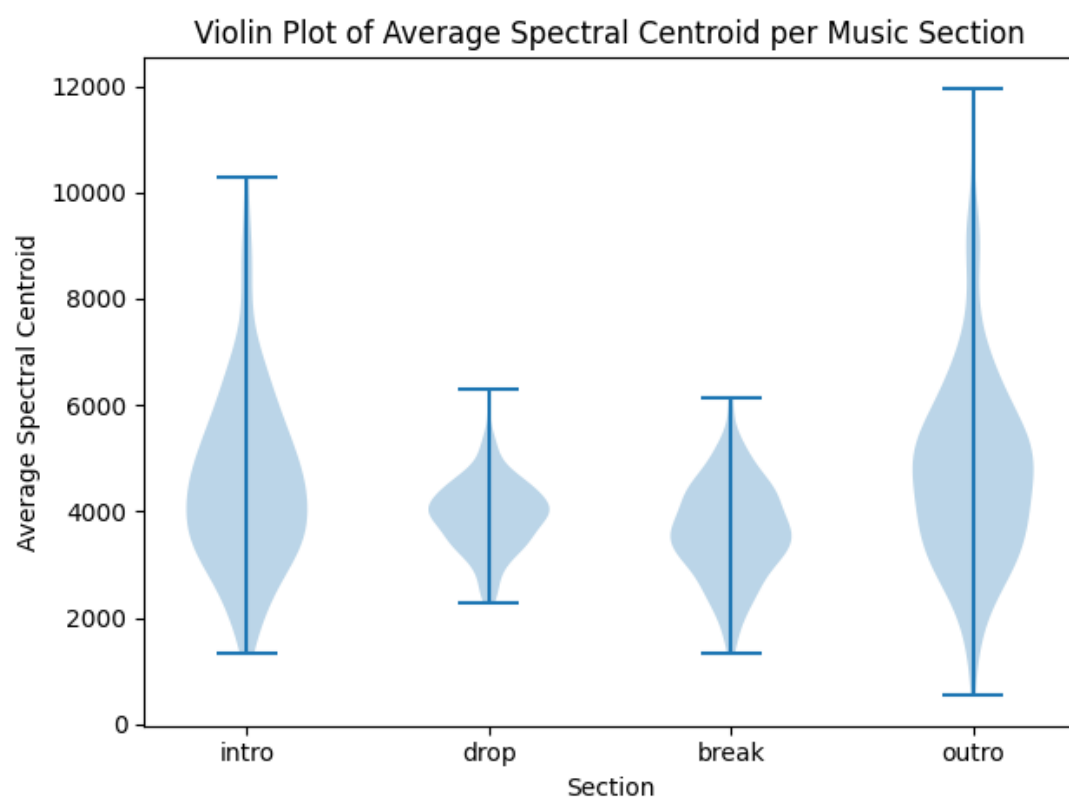


図 5.2: セクション別の Spectral Centroid の平均値からなるバイオリングラフ

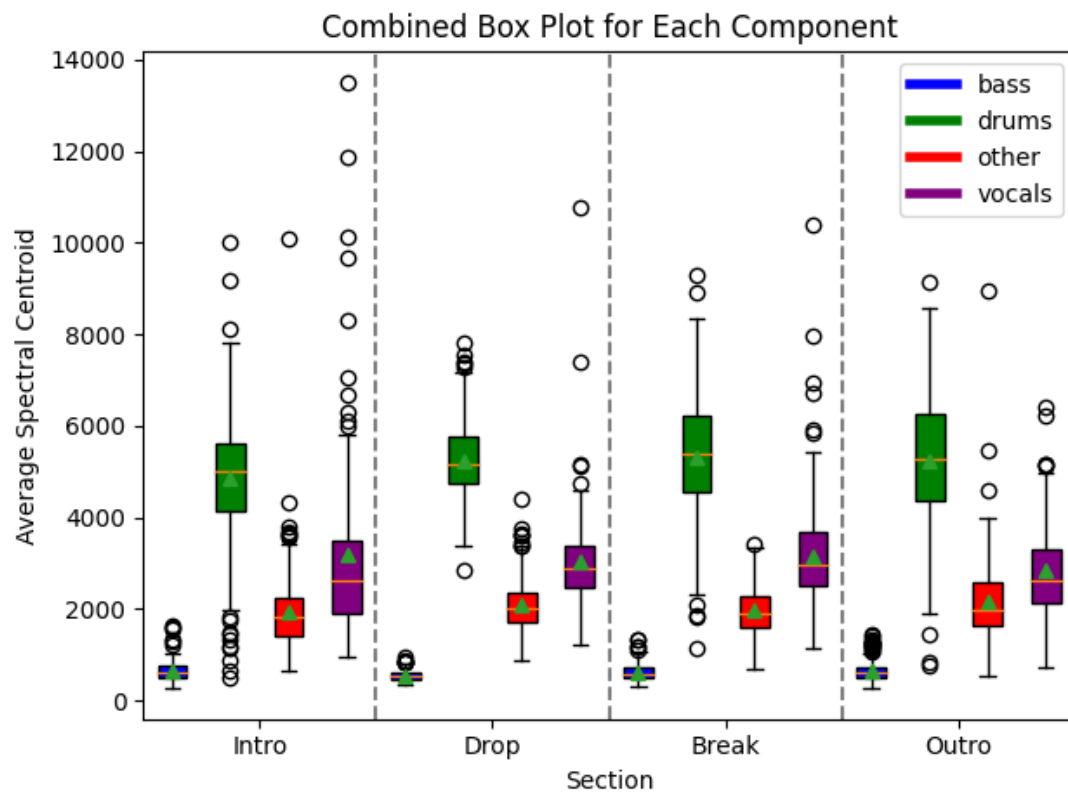


図 5.3: 各パートにおけるセクション別の Spectral Centroid の平均値からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. 緑の三角は平均値を表す. オレンジの線は中央値を表す.

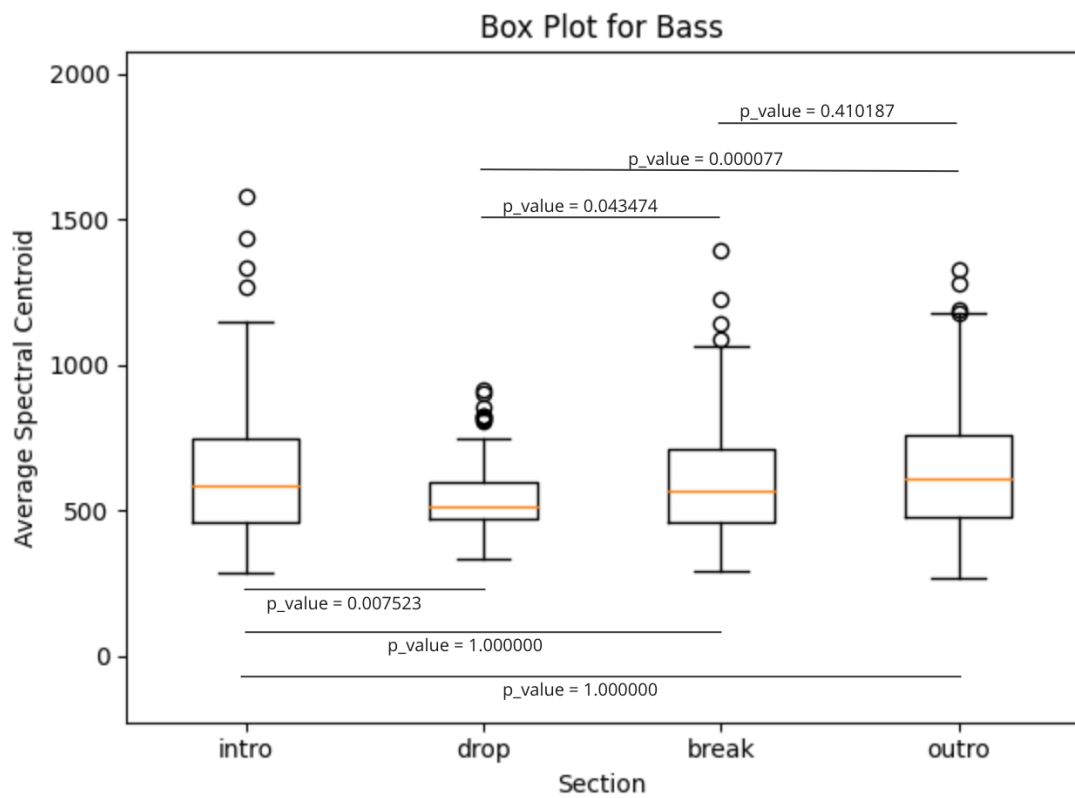


図 5.4: ベースにおけるセクション別の Spectral Centroid の平均値からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレンジの線は中央値を表す.

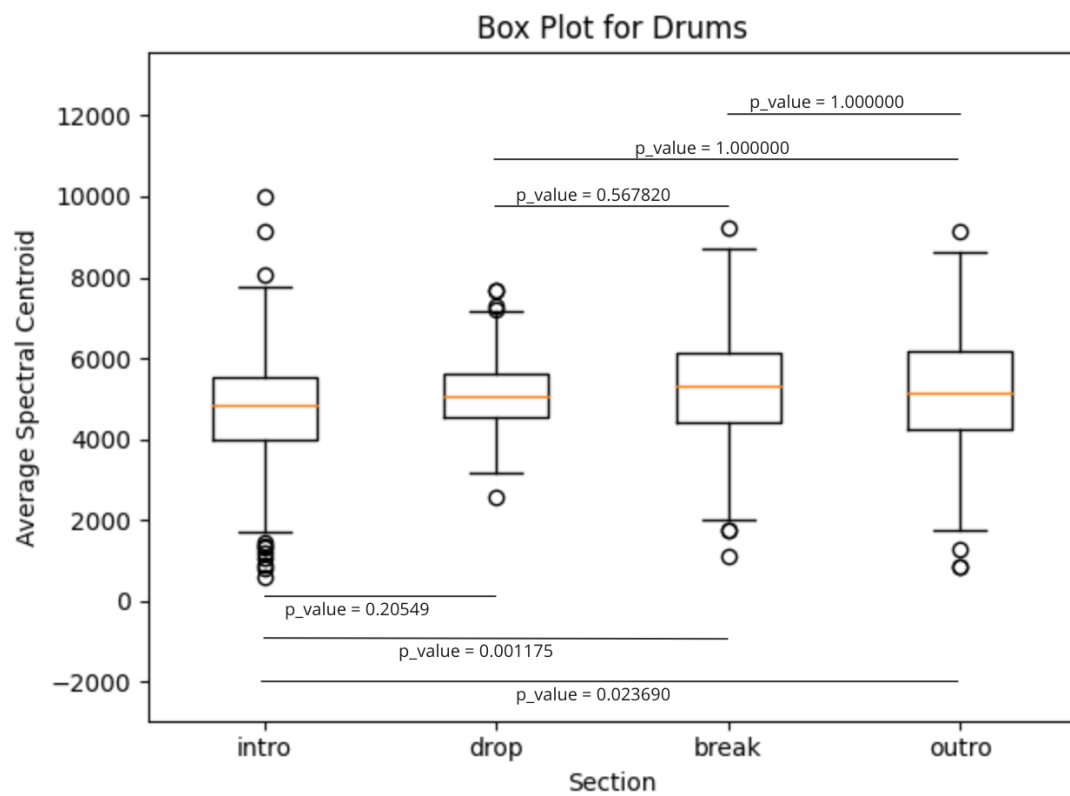


図 5.5: ドラムにおけるセクション別の Spectral Centroid の平均値からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレンジの線は中央値を表す.

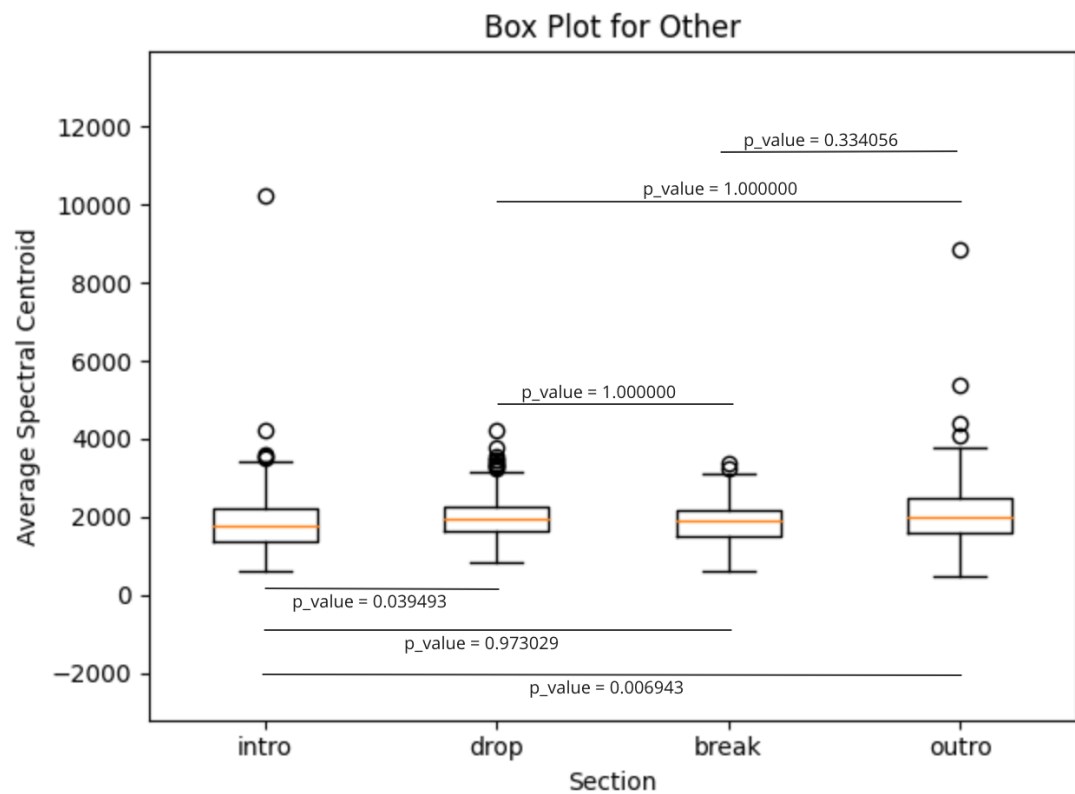


図 5.6: その他のパートにおけるセクション別の Spectral Centroid の平均値からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレンジの線は中央値を表す.

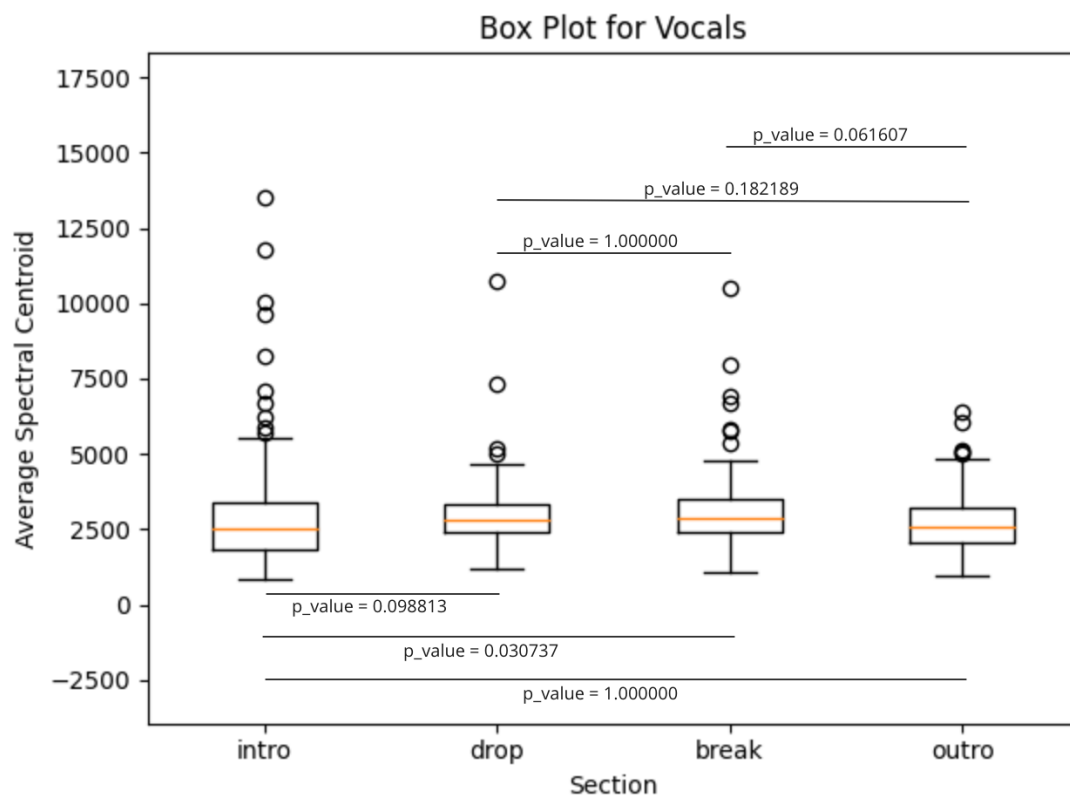


図 5.7: ヴォーカルにおけるセクション別の Spectral Centroid の平均値からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレンジの線は中央値を表す.

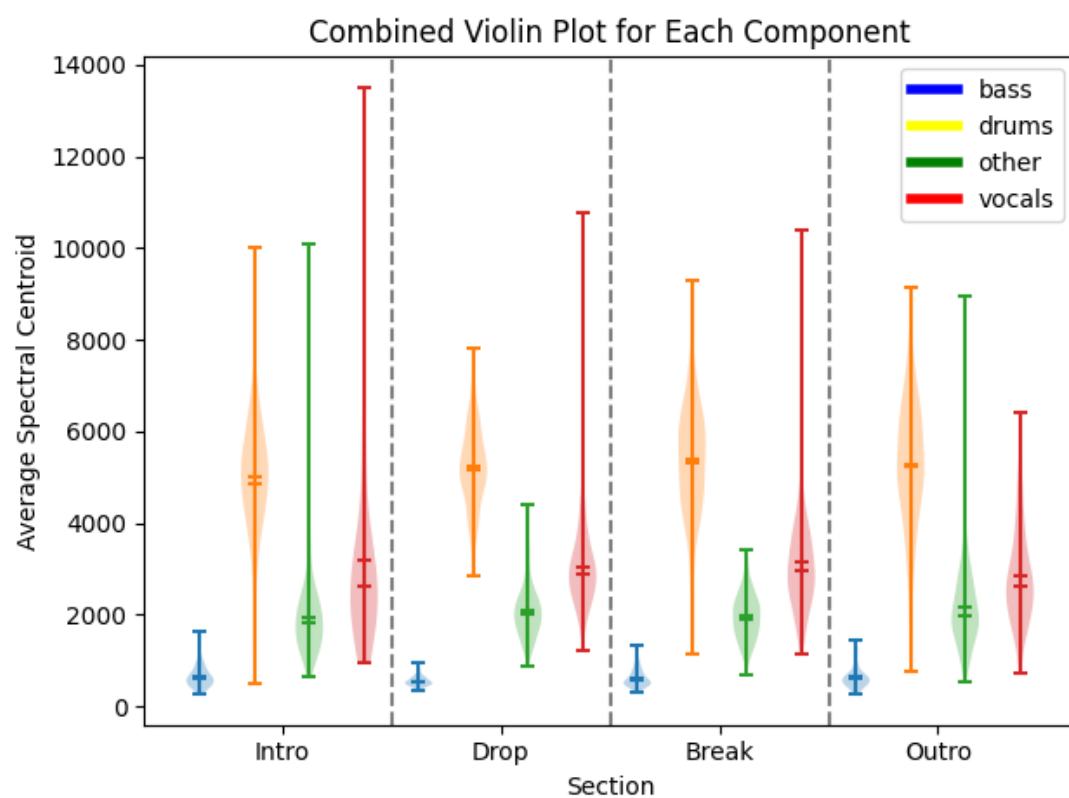


図 5.8: 各パートにおけるセクション別の Spectral Centroid の平均値からなるバイオリングラフ

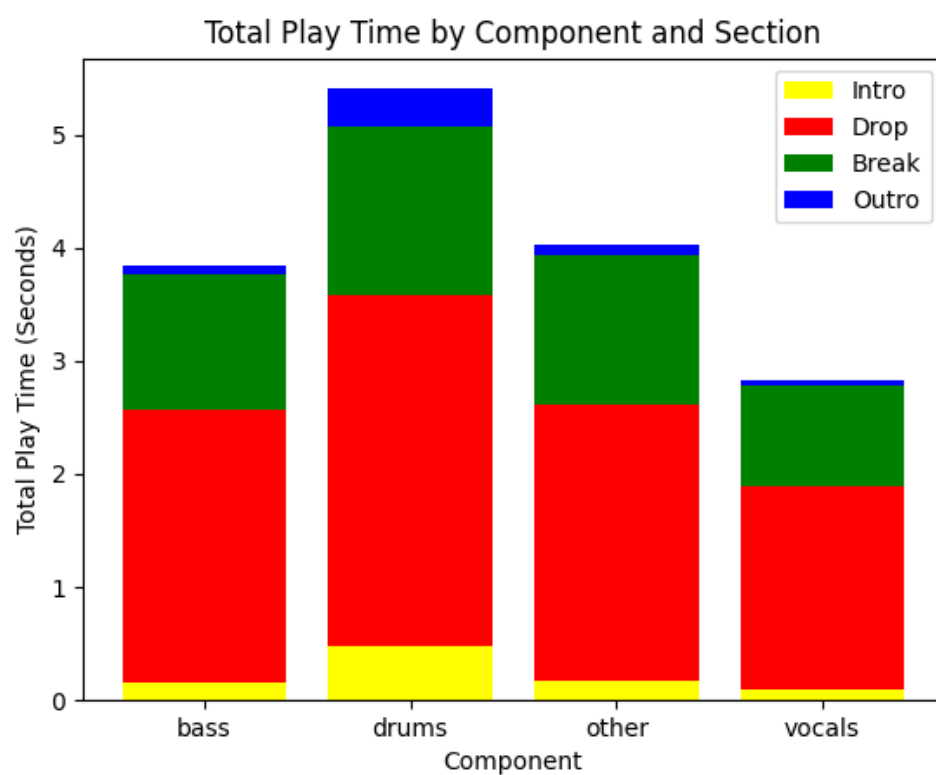


図 5.9: RMS が閾値を下回る区間を除外した後の各パートのセクション別の有効区間



表 5.4: ベースのセクション別のダンの多重比較検定の結果 (p 値)

ベース	イントロ	ドロップ	ブレイク	アウトロ
イントロ	1.000	$6.655 \times 10^{-15}$	$8.451 \times 10^{-18}$	$2.344 \times 10^{-14}$
ドロップ	$6.655 \times 10^{-15}$	1.000	$8.127 \times 10^{-64}$	$3.260 \times 10^{-57}$
ブレイク	$8.451 \times 10^{-18}$	$8.127 \times 10^{-64}$	1.000	1.000
アウトロ	$2.344 \times 10^{-14}$	$3.260 \times 10^{-57}$	1.000	1.000

表 5.5: ドラムのセクション別のダンの多重比較検定の結果 (p 値)

ドラム	イントロ	ドロップ	ブレイク	アウトロ
イントロ	1.000	$1.161 \times 10^{-43}$	0.097042	1.000
ドロップ	$1.161 \times 10^{-43}$	1.000	$1.443 \times 10^{-30}$	$6.114 \times 10^{-49}$
ブレイク	0.097042	$1.443 \times 10^{-30}$	1.000	0.010997
アウトロ	1.000	$6.114 \times 10^{-49}$	0.010997	1.000

### 5.1.2 音圧に関する分析

分析結果を図 5.10～図 5.19, 表 5.7～表 5.14 に示す.

図 5.10, 図 5.11 が示している通り, 各セクションの音圧には大きな差がある. このことは, 表 5.4 の検定結果が裏付けている. ドロップが最も高く, 次いでブレイク, イントロ, アウトロの順に高い. また, ドロップとブレイクは音圧が中央値または平均値付近を中心に値が集中している一方で, イントロとアウトロは値の分散が大きい. 予備実験ではドロップとイントロ・アウトロの音圧に差が無い場合もあったが, 実際には, イントロとアウトロが取りうる音圧には幅があり, 大半はドロップの方が音圧が高いということが分かった.

パート別の可視化である図 5.12, 図 5.17 を見ると, 予備実験の結果と同様に, ドロップではドラムとベースの音圧が他のパートと比較して高く, ブレイクではこれらの音圧が低下している. また, イントロとアウトロではドラムの音圧が他

表 5.6: その他のパートのセクション別のダンの多重比較検定の結果 (p 値)

その他	イントロ	ドロップ	ブレイク	アウトロ
イントロ	1.000	1.000	$6.750 \times 10^{-39}$	$4.973 \times 10^{-42}$
ドロップ	1.000	1.000	$6.672 \times 10^{-35}$	$6.789 \times 10^{-38}$
ブレイク	$6.750 \times 10^{-39}$	$6.672 \times 10^{-35}$	1.000	1.000
アウトロ	$4.973 \times 10^{-42}$	$6.789 \times 10^{-38}$	1.000	1.000

表 5.7: ヴォーカルのセクション別のダンの多重比較検定の結果 (p 値)

ヴォーカル	イントロ	ドロップ	ブレイク	アウトロ
イントロ	1.000	0.6006	$5.111 \times 10^{-41}$	$3.187 \times 10^{-41}$
ドロップ	0.6006	1.000	$1.147 \times 10^{-32}$	$7.990 \times 10^{-33}$
ブレイク	$5.111 \times 10^{-41}$	$1.147 \times 10^{-32}$	1.000	1.000
アウトロ	$3.187 \times 10^{-41}$	$7.990 \times 10^{-33}$	1.000	1.000

のパートと比較して高い。ベースに関してはドロップ以外のセクションでは音圧が低いので、ベースの音圧が高い場合はドロップである可能性が高いという見方が出来る。ヴォーカルとその他のパートは、ドロップとブレイクでは大きな差が無く、イントロとアウトロではほとんど音が鳴っていないことが分かる。このことから、ヴォーカルとその他の音圧が高い場合は、ドロップもしくはブレイクである可能性が高いことを示唆している。

3パートの音圧バランスをセクション別に可視化した図 5.18, 図 5.19 では、ドロップではドラムの音圧が全セクションの中で最も高い位置に分布しており、ベースとヴォーカル+その他のパートの音圧も比較的高い位置に分布している。ブレイクではヴォーカル+その他のパートの音圧はドロップと同程度であるが、ドラムとベースの音圧はドロップより低い。イントロではドラムの音圧に幅があるものの、ベースとヴォーカル+その他のパートはほとんど音が鳴っていない。これらの傾向

は図 5.12, 図 5.17 で確認されたものと一致している.

一方, イントロとアウトロに関しては, アウトロではドラムとベースの音圧はイントロと変わらず, しかし, ヴォーカル+その他のパートの音圧はイントロと比較して高くなっている. これは, ハウスミュージックはイントロがドラムのみで始まり, アウトロがドラムとヴォーカルまたはその他のパートで終わるパターンが一般的であることと示唆している.

表 5.8: セクション別のクラスカル・ウォリス検定結果

統計量	400.5070567814819
p 値	$1.719 \times 10^{-86}$

表 5.9: セクション別のダンの多重比較検定の結果 (p 値)

	イントロ	ドロップ	ブレイク	アウトロ
ブレイク	1.000	$9.729 \times 10^{-27}$	0.002876	$3.356 \times 10^{-15}$
ドロップ	$9.729 \times 10^{-27}$	1.000	$1.020 \times 10^{-46}$	$3.167 \times 10^{-81}$
イントロ	0.002876	$1.020 \times 10^{-46}$	1.000	$2.264 \times 10^{-5}$
アウトロ	$3.356 \times 10^{-15}$	$3.167 \times 10^{-81}$	$2.264 \times 10^{-5}$	1.000

表 5.10: 各パートのセクション別のクラスカル・ウォリス検定結果

パート	統計量	p 値
ベース	377.15480136069164	$1.964 \times 10^{-81}$
ドラム	285.1885099565866	$1.596 \times 10^{-61}$
その他	343.55226926076966	$3.713 \times 10^{-74}$
ボーカル	330.4136558583632	$2.596 \times 10^{-71}$

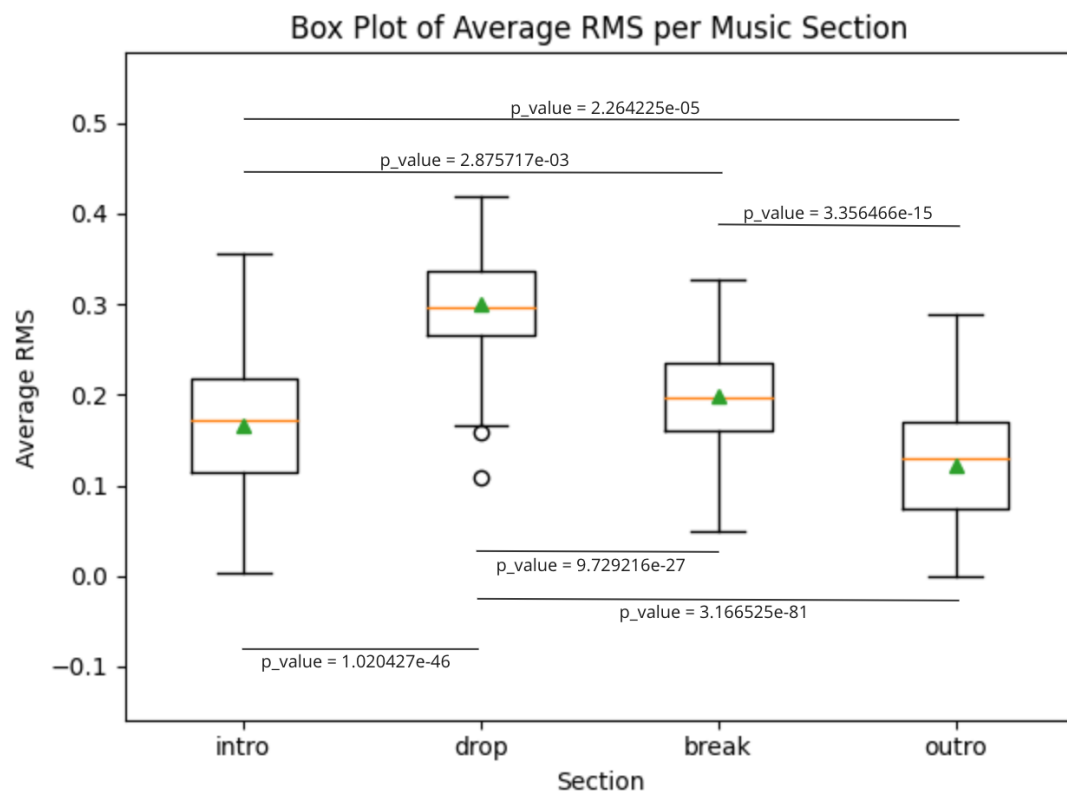


図 5.10: セクション別の RMS の平均値からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. 緑の三角は平均値を表す. オレンジの線は中央値を表す.

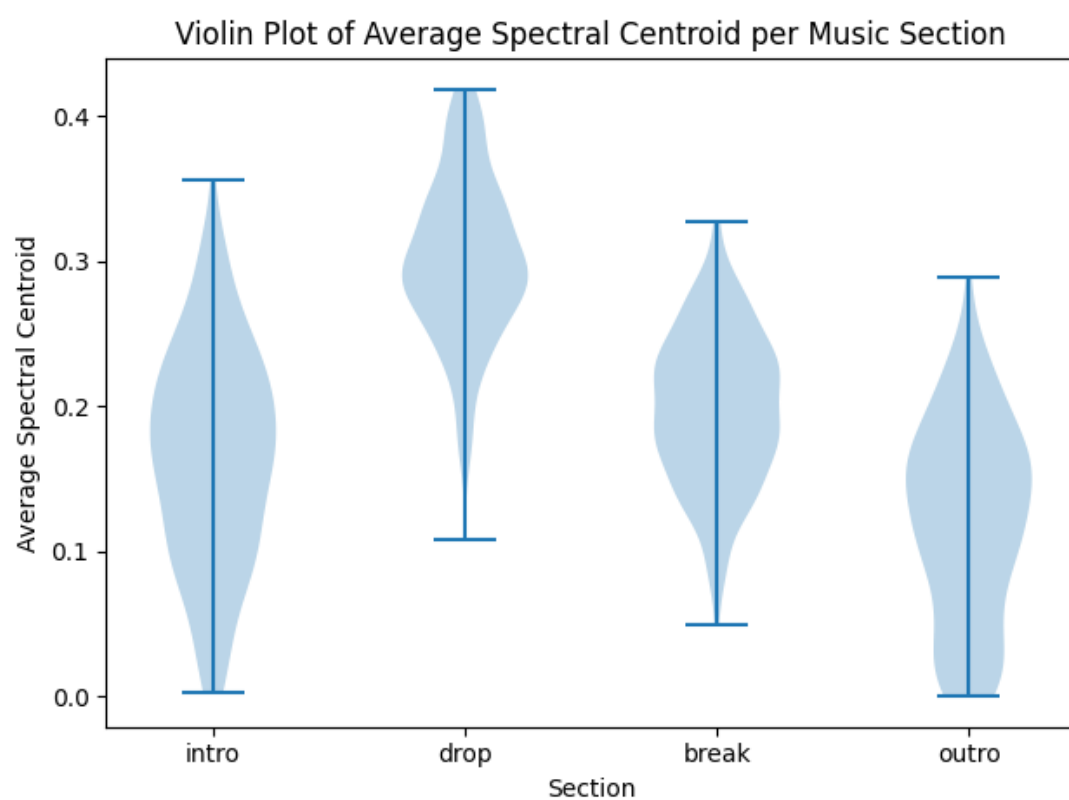


図 5.11: セクション別の RMS の平均値からなるバイオリングラフ

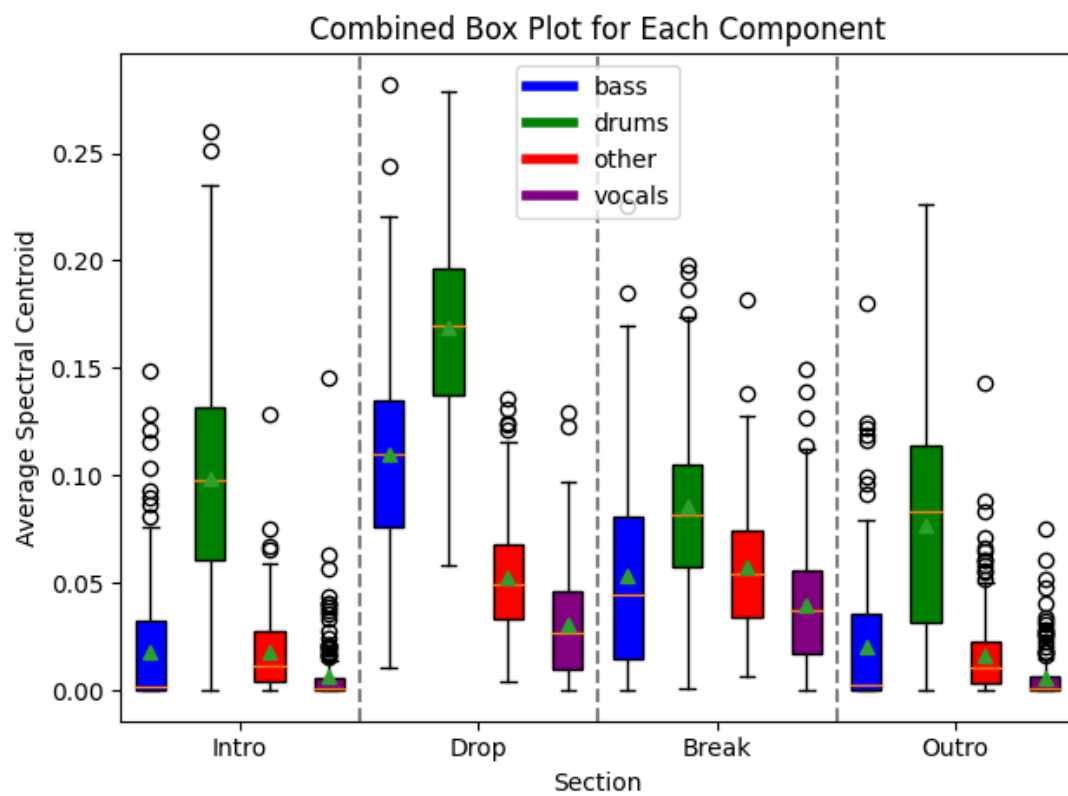


図 5.12: 各パートにおけるセクション別の RMS の平均値からなる箱ひげ図。白丸は四分位範囲の 1.5 倍を超えた外れ値を表す。緑の三角は平均値を表す。オレンジの線は中央値を表す。

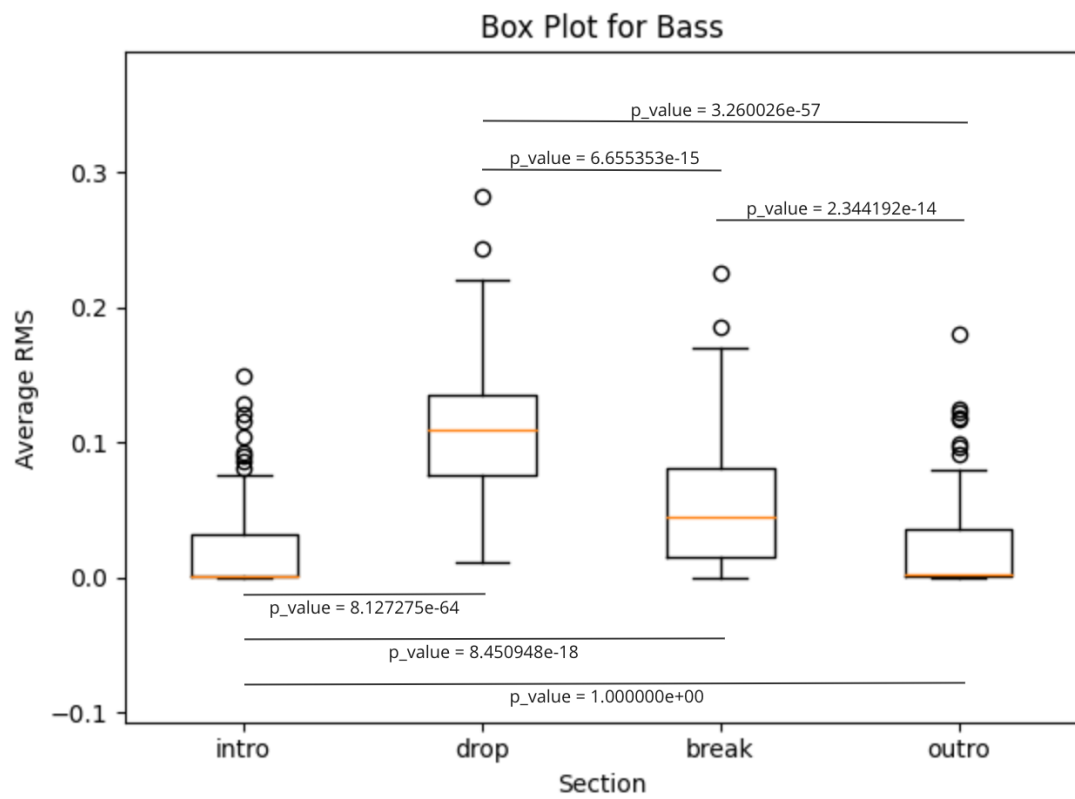


図 5.13: ベースにおけるセクション別の RMS の平均値からなる箱ひげ図。白丸は四分位範囲の 1.5 倍を超えた外れ値を表す。オレンジの線は中央値を表す。

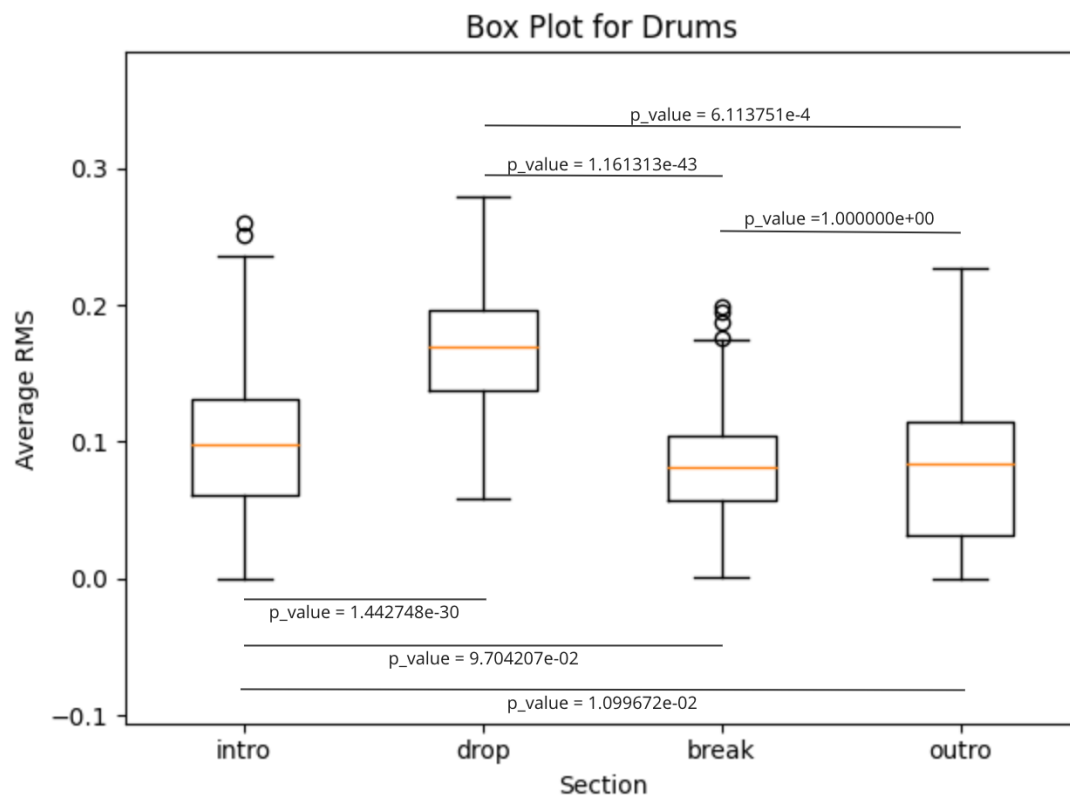


図 5.14: ドラムにおけるセクション別の RMS の平均値からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレンジの線は中央値を表す.



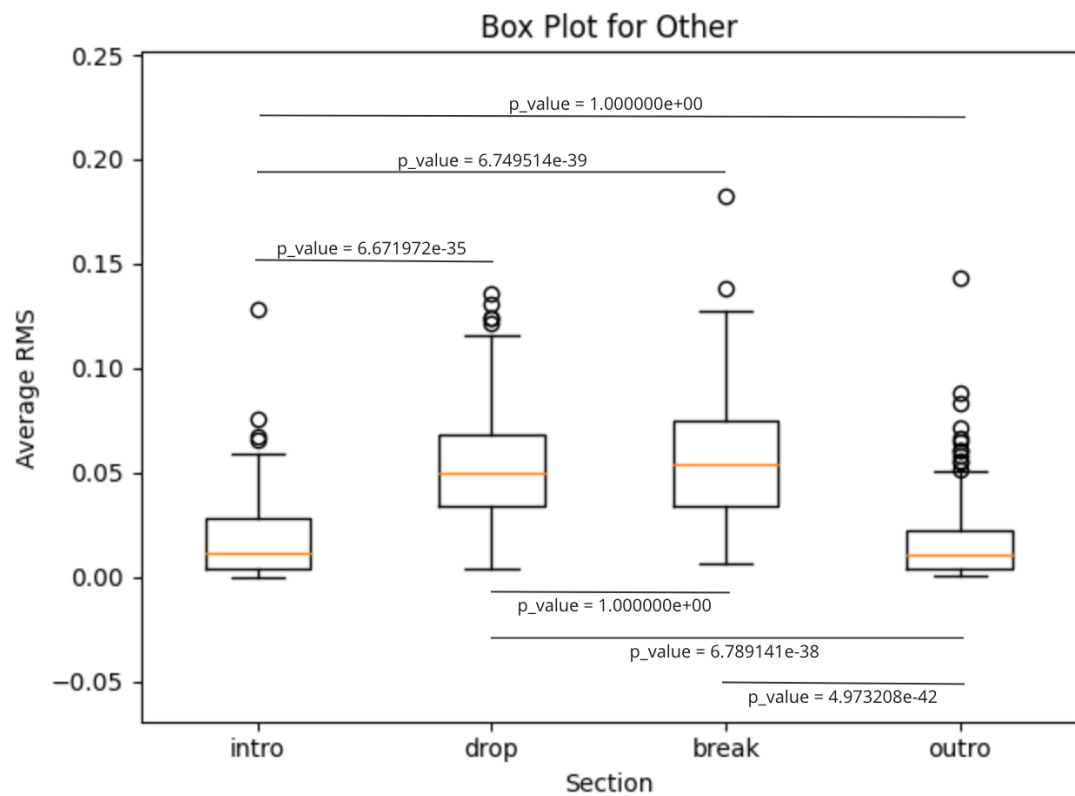


図 5.15: その他のパートにおけるセクション別の RMS の平均値からなる箱ひげ図. 白丸は四分位範囲の 1.5 倍を超えた外れ値を表す. オレンジの線は中央値を表す.

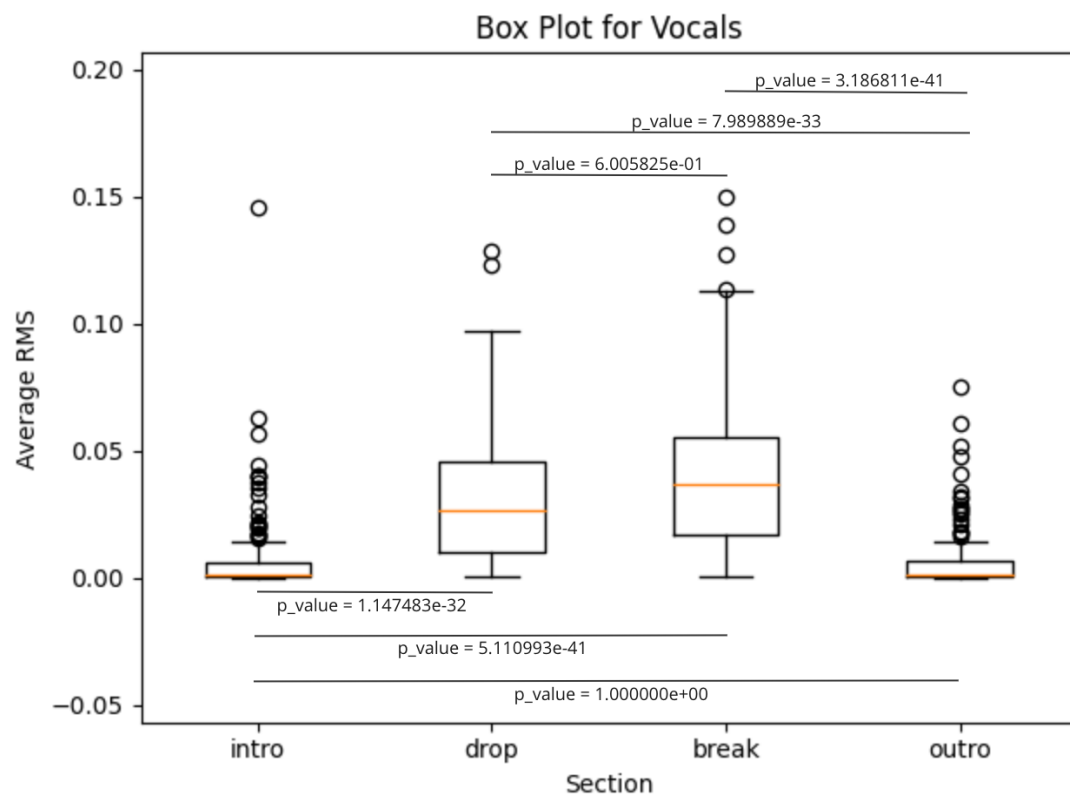


図 5.16: ヴォーカルにおけるセクション別の RMS の平均値からなる箱ひげ図。白丸は四分位範囲の 1.5 倍を超えた外れ値を表す。オレンジの線は中央値を表す。

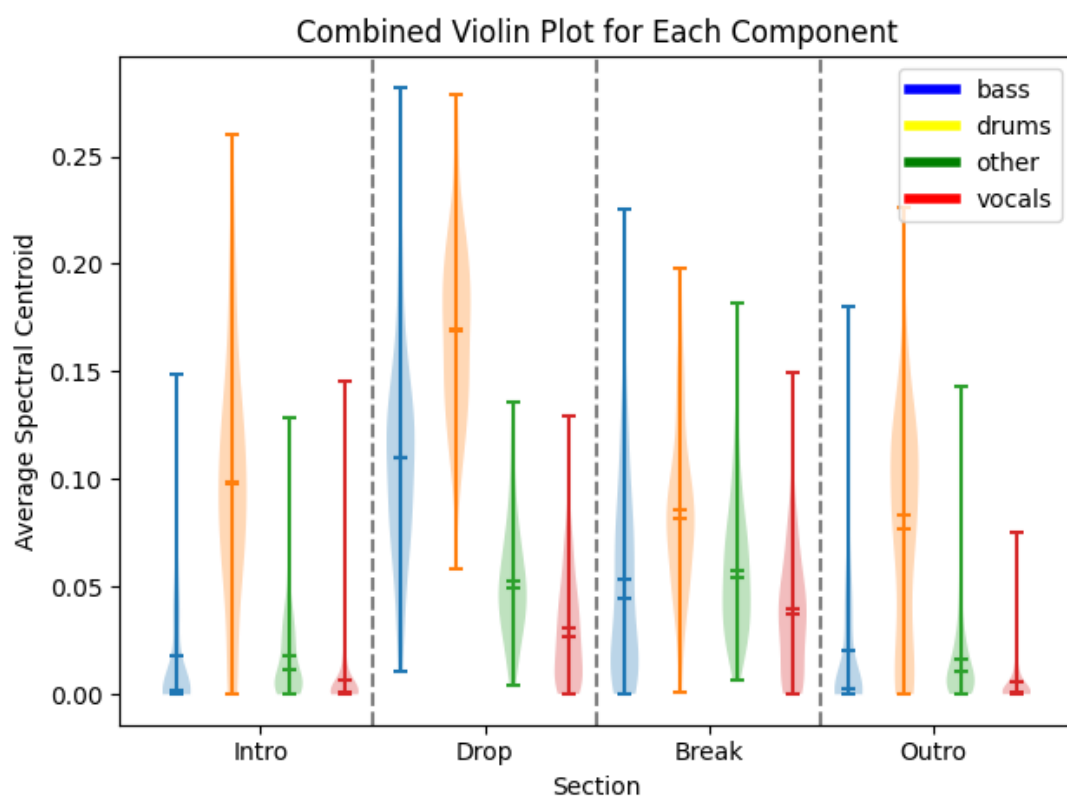


図 5.17: 各パートにおけるセクション別の RMS の平均値からなるバイオリングラフ

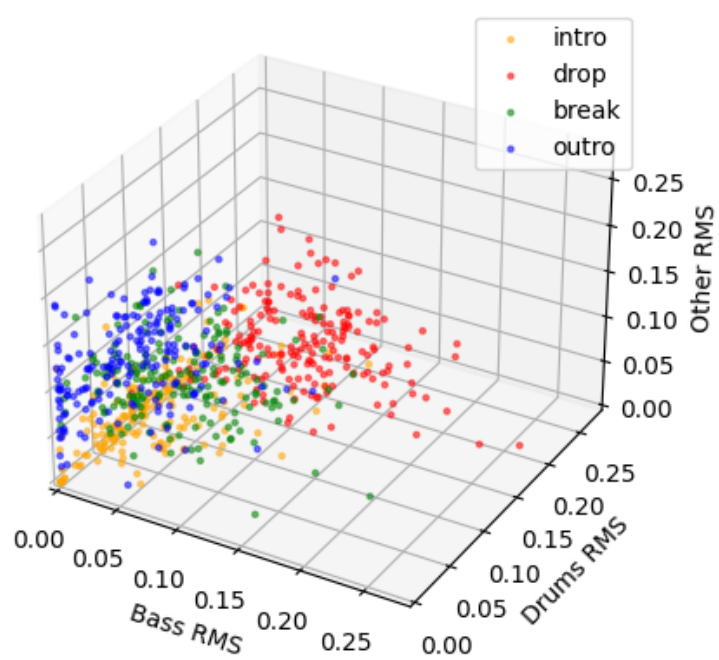


図 5.18: セクション別に点を色分けして3パートの音圧バランスを3次元座標空間にプロットしたもの

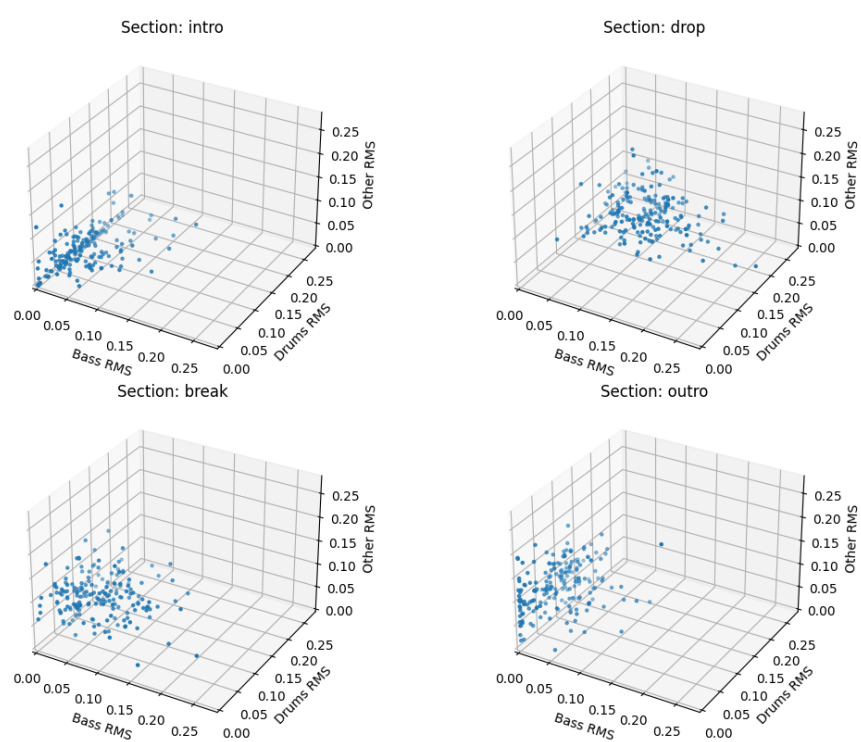


図 5.19: 図 5.18 の図についてセクション別に座標空間を分けたもの

表 5.11: ベースのセクション別のダンの多重比較検定の結果 (p 値)

ベース	イントロ	ドロップ	ブレイク	アウトロ
イントロ	1.000	$6.655 \times 10^{-15}$	$8.451 \times 10^{-18}$	$2.344 \times 10^{-14}$
ドロップ	$6.655 \times 10^{-15}$	1.000	$8.127 \times 10^{-64}$	$3.260 \times 10^{-57}$
ブレイク	$8.451 \times 10^{-18}$	$8.127 \times 10^{-64}$	1.000	1.000
アウトロ	$2.344 \times 10^{-14}$	$3.260 \times 10^{-57}$	1.000	1.000

表 5.12: ドラムのセクション別のダンの多重比較検定の結果 (p 値)

ドラム	イントロ	ドロップ	ブレイク	アウトロ
イントロ	1.000	$1.161 \times 10^{-43}$	0.097042	1.000
ドロップ	$1.161 \times 10^{-43}$	1.000	$1.443 \times 10^{-30}$	$6.114 \times 10^{-49}$
ブレイク	0.097042	$1.443 \times 10^{-30}$	1.000	0.010997
アウトロ	1.000	$6.114 \times 10^{-49}$	0.010997	1.000

### 5.1.3 ドラム音に関する分析

分析結果を図 5.20～図 5.27 に示す。

セクションごとのドラム要素の個数の表す図 5.20 が示している通り、セクションごとにおける各ドラム要素の発音回数割合に大きさ差はない。このことから、セクションの違いがドラム要素の種類に変化をもたらすことはないと言える。

また、8 小節あたりの各ドラム要素の発音回数を表す図 5.21, 図 5.22, 図 5.23 では、キックは 8 小節あたり 64 個程度、スネアは 8 小節あたり 32 個程度、ハイハットは 8 小節あたり 128 個程度の発音回数に集中していることが見れ取れる。キックは 8 分音符、スネアは 4 分音符、ハイハットは 16 分音符で演奏されていることを示している。また、キックと比較してスネアとハイハットの分散が大きく、これらは変化しやすい要素であると考えられる。反対に、キックの分散が少ないことから、どの楽曲でも一定のリズムを保ちやすいと言える。ただし、いずれのドラ

表 5.13: その他のパートのセクション別のダンの多重比較検定の結果 (p 値)

その他	イントロ	ドロップ	ブレイク	アウトロ
イントロ	1.000	1.000	$6.750 \times 10^{-39}$	$4.973 \times 10^{-42}$
ドロップ	1.000	1.000	$6.672 \times 10^{-35}$	$6.789 \times 10^{-38}$
ブレイク	$6.750 \times 10^{-39}$	$6.672 \times 10^{-35}$	1.000	1.000
アウトロ	$4.973 \times 10^{-42}$	$6.789 \times 10^{-38}$	1.000	1.000

表 5.14: ヴォーカルのセクション別のダンの多重比較検定の結果 (p 値)

ヴォーカル	イントロ	ドロップ	ブレイク	アウトロ
イントロ	1.000	0.6006	$5.111 \times 10^{-41}$	$3.187 \times 10^{-41}$
ドロップ	0.6006	1.000	$1.147 \times 10^{-32}$	$7.990 \times 10^{-33}$
ブレイク	$5.111 \times 10^{-41}$	$1.147 \times 10^{-32}$	1.000	1.000
アウトロ	$3.187 \times 10^{-41}$	$7.990 \times 10^{-33}$	1.000	1.000

ム要素も、8小節あたりの発音回数が0になる部分があり、これはドラムパターンの変化に起因すると考えられる。また、いずれのドラム要素に関しても、時間軸に沿った傾向の変化は見られず、ハウスミュージックにおけるドラムパターンの変化は楽曲によってタイミングが異なると言える。

さらに、ドラムパターンとセクションの変化点の一致率を示す図 5.24, 図 5.25 では、一致率の最頻値は30%から40%程度であることが読み取れる。また、表 5.15 と表 5.15 から、それぞれ一致率の平均値は約36%であり、およそ3回に1回の頻度でドラムパターンの変化とセクションの変化が同時に発生していることになる。このことから、ドラムパターンの変化とセクションの変化は相関関係にあると言える。

最後に、一致が発生した時間帯を示す図 5.26 と図 5.27 では、楽曲の開始地点を除いて、どの時間帯でも一致していることが分かる。これは、ドラムパターン

とセクションの一致が時間軸とは関連なく発生していることを示唆している。また，楽曲の開始時点で一致回数が増加しているのは，ドラムパターンの検出アルゴリズムに起因するものと考えられる。

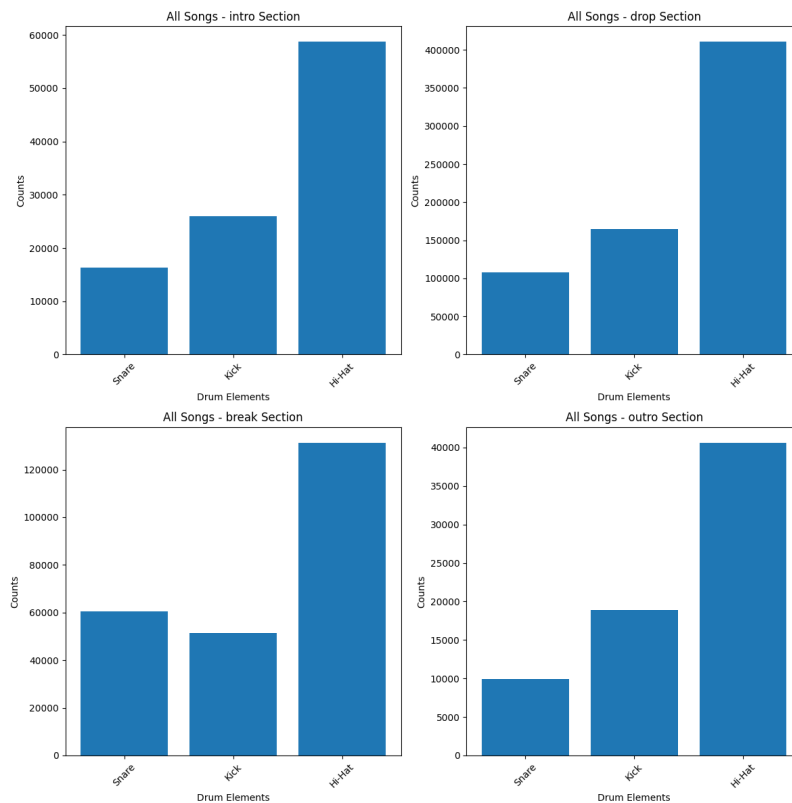


図 5.20: 各セクションにおける各ドラム要素の発音回数からなる棒グラフ

表 5.15: ドラムパターンの変化点を分母とした場合の一致率の平均値

平均値	36.78%
-----	--------

#### 5.1.4 総合考察

考察から，周波数・音圧・ドラム音の3つの要素とハウスミュージックの楽曲構成の関連性について，セクション間での変化が無い要素とある要素の2つに分類



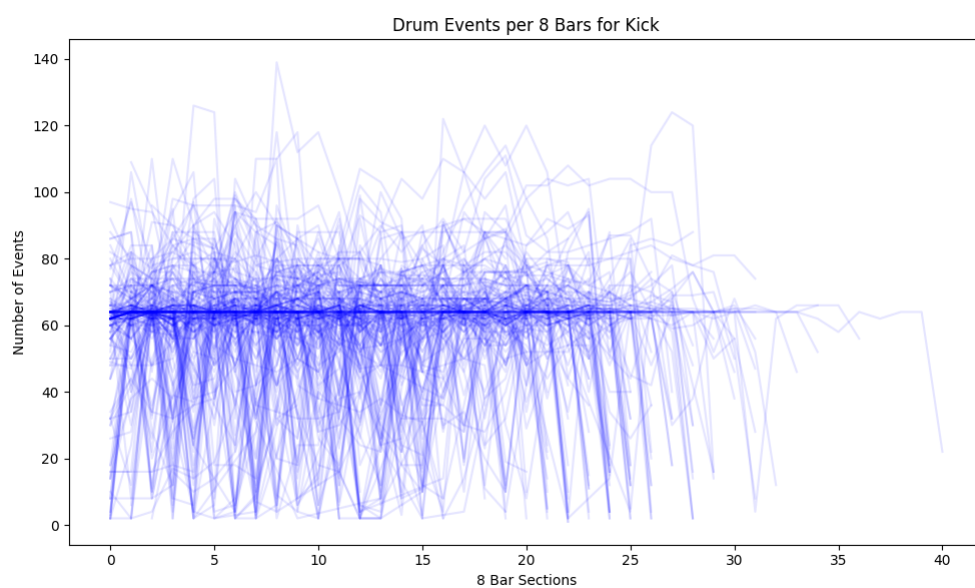


図 5.21: キックの8小節ごとの発音回数を折れ線グラフとしたスパゲッティプロット

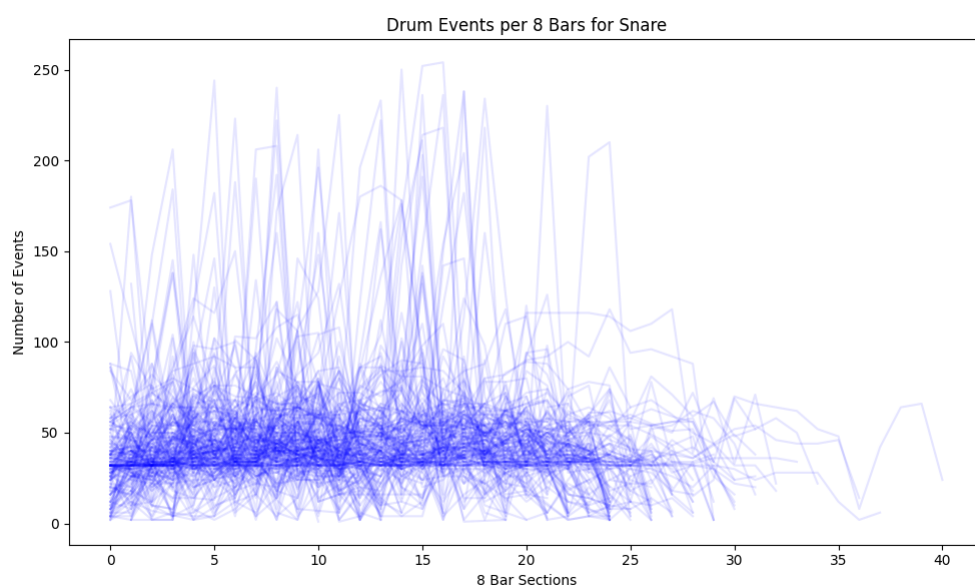


図 5.22: スネアの8小節ごとの発音回数を折れ線グラフとしたスパゲッティプロット

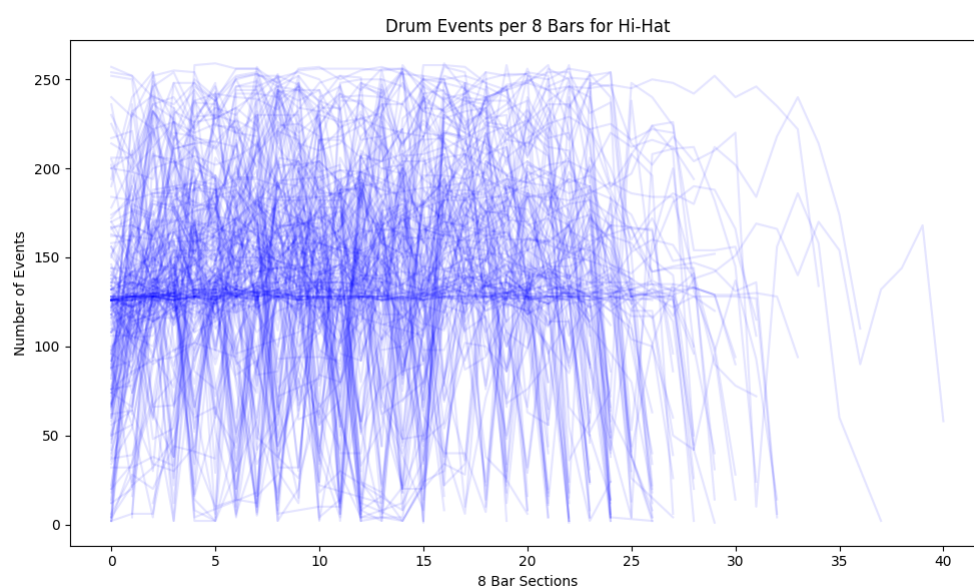


図 5.23: ハイハットの8小節ごとの発音回数を折れ線グラフとしたスパゲッティプロット

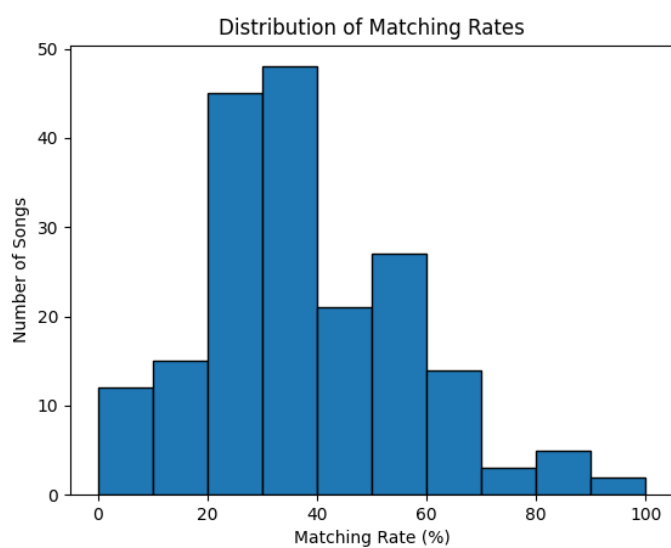


図 5.24: ドラムパターンの変化点を分母とした場合の一致率のヒストグラム

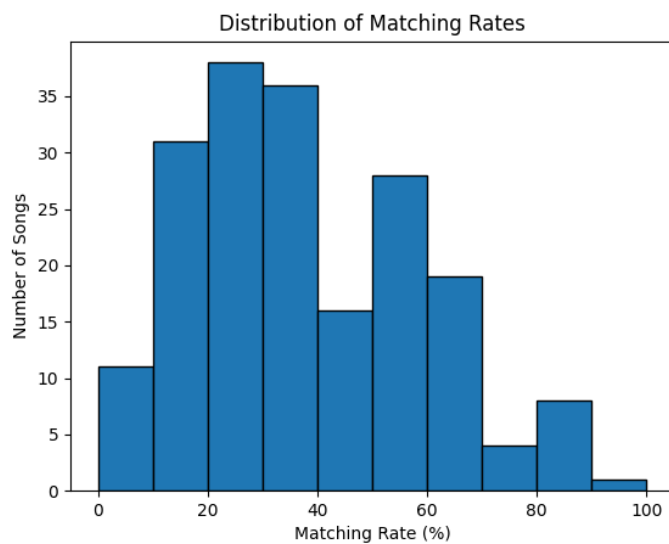


図 5.25: セクションの変化点を分母とした場合の一致率のヒストグラム

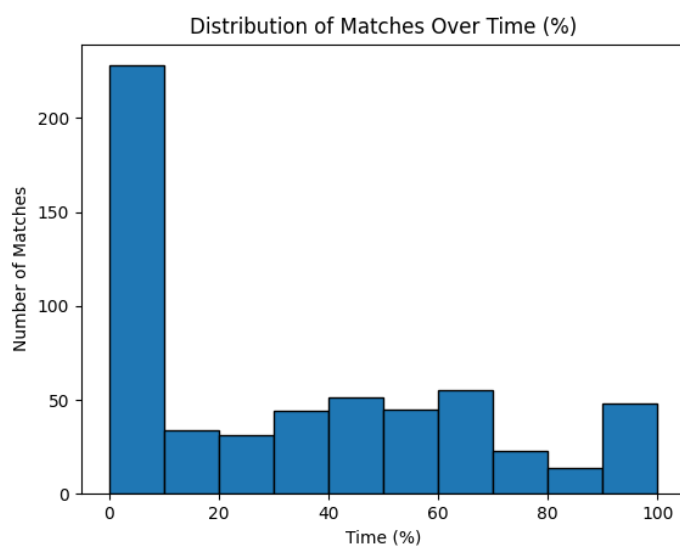


図 5.26: ドラムパターンの変化点を分母とした場合の一致が発生した楽曲における時間帯のヒストグラム

表 5.16: セクションの変化点を分母とした場合の一致率の平均値

平均値	36.68%
-----	--------

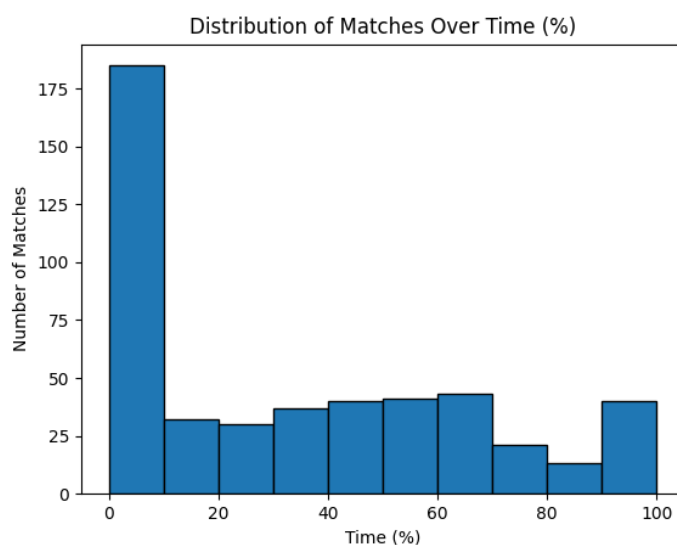


図 5.27: セクションの変化点を分母とした場合の一致が発生した楽曲における時間帯のヒストグラム

することが出来る。

セクション間で変化がない要素としては、周波数、各ドラム要素、ドラムパターンが挙げられる。例えば周波数については、各セクションにおける Spectral Centroid の中央値または平均値は、いずれも 4000Hz から 4500Hz 程度に収まっており、大きな差は無かった。また、各ドラム要素では、各セクションにおけるそれぞれのドラム要素の発音回数の割合は、異なるセクションでも変化が無かった。更に、ドラムパターンとセクションの変化点の一致位置は時間軸に相関が無かった。これらの要素はセクションに関わらず一定しており、楽曲全体の安定性や統一感をもたらしている。

一方、セクション間で変化がある要素としては、音圧、そして鳴っているパー

トの違いが挙げられる。例えば音圧については、各セクションの音圧に差があり、ドロップ、ブレイク、イントロ、アウトロの順番に音圧が高かった。また、鳴っているパートの違いとして、イントロではヴォーカルまたはその他のパートが殆ど鳴っていないのに対して、アウトロではベースが殆ど鳴っていなかった。これらの要素はセクションごとに異なり、楽曲に起承転結やセクションの変化をもたらしている。

また、セクションの変化と相関関係にあるドラムパターンの変化は、これらの要素によって生じるセクションの変化を強調する役割を持っていると考えられる。このことから、ハウスミュージックではセクションごとの特徴を明確に区分しつつも、一定の特徴は維持することで楽曲としての統一感を保っていると考えられる。このバランスの取り方が、ハウスミュージックの楽曲構成の特徴の一つと言えるだろう。

## 第6章 まとめ

本章では、実験の結論と、本研究の今後の展望について述べる。

### 6.1 結論

本稿では、ハウスミュージックの楽曲構成について、それを決定する要因や、楽曲構成における法則性を調査することを目的として実験を行った。関連研究では楽曲構成の可視化手法が多く提案されているが、ハウスミュージック特有の分析手法の提案は無かった。そこで、本稿ではハウスミュージックの特徴を踏まえた分析手法を用い、楽曲構成の分析を行った。予備実験では、ハウスミュージック特有の特徴である各パートの重なりを基に、音圧と楽曲構成の関連性について調べる実験を行った、その結果から、周波数やドラム音とハウスミュージックの楽曲構成の関連性について更に深く調査することにした。

本研究では、All-In-One を用いて楽曲のセグメンテーション情報を取得し、セクションごとの周波数特徴量やRMSの違い、各ドラム要素の発音回数の違いなどを分析した。実験の結果、周波数に関しては、セクション間で有意な周波数帯の変化は見られず、各パートが一定の周波数帯を保持していることが明らかになった。これは、周波数の変化がセクションの変化を引き起こすわけではないことを示している。また、ドロップとブレイクでは Spectral Centroid の値が 4000Hz 付近に集中しており、イントロとアウトロでは Spectral Centroid の分散が大きく、より多様な音色が使われていることが分かった。音圧に関しては、ドロップが最も音圧が高く、次いでブレイク、イントロ、アウトロの順に音圧が高いことを確

認した。これはドロップがハウスミュージックにおいて最もエネルギーが高まるセクションであることを裏付けている。また、イントロとアウトロの違いとして、アウトロではその他もしくはヴォーカルセクションが鳴りやすいことが分かった。ドラム音に関しては、セクションによるドラム要素の個数に有意差はなく、セクションによるドラムパターンの音色に変化は無いことがわかった。また、キックは楽曲全体を通して一定を保っており、ハウスミュージックのリズムの基盤としての役割が強調されており、一方で、スネアとハイハットはより多様なパターンを示し、楽曲内での変化に富んでいることが分かった。

これらの結果から、ハウスミュージックの楽曲構成では、セクションごとに明確に区分される特徴と、楽曲全体としての統一感を保つ要素が共存しており、このバランスの取り方が、ハウスミュージックの楽曲構成の特徴の一つだと考えられることが分かった。

## 6.2 今後の展望・課題

本稿で行われた実験の課題として、まず All-In-One の精度の問題が挙げられる。All-In-One で得られるセグメンテーション情報には、主観的に見て一定数の誤りが存在しているように感じられた。また、ハウスミュージックのセクションの定義自体が業界で統一されていないため、本研究で用いた4つのセクションの妥当性については、有識者の評価を通じて検証する必要がある。

次に、本研究では194曲の楽曲データを使用した。ハウスミュージックは歴史が長く多岐にわたるジャンルであるため、使用した楽曲データがハウスミュージック全体を完全に網羅しているとは言い切れない。このことから、更に実験データを増やして実験を行う必要がある。

また、ドラム音の分析におけるドラムパターンの変化検出アルゴリズムは、ドラムイベント間の間隔の長さを基に変化を検出するため、細かなドラムパターン

の変化を捉えきれないという問題があった。このアルゴリズムの改善も今後の課題である。

さらに、本研究はハウスミュージックに限定して実施されたが、今後はテクノやディスコなど他のダンスミュージックジャンルとの比較分析を行うことで、ハウスミュージックの特徴をより客観的に把握することが出来ると考えられる。





## 参考文献

- [1] Jonathan Foote: Visualizing Music and Audio using Self-Similarity, Proceedings of the Seventh ACM International Conference on Multimedia (Part 1), pp. 77—80 , 1999.
- [2] Jouni Paulus, and Anssi Klapuri: Music Structure Analysis Using a Probabilistic Fitness Measure and a Greedy Search Algorithm, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 17, Issue: 6, pp. 1159–1170 , 2009.
- [3] Taejun Kim, Juhan Nam: All-In-One Metrical And Functional Structure Analysis With Neighborhood Attentions on Demixed Audio: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2023.
- [4] J. Salamon, Oriol Nieto, Nicholas J. Bryan: Deep Embeddings and Section Fusion Improve Music Segmentation, in ISMIR , 2021.
- [5] Brian McFee, Oriol Nieto, Juan Pablo Bello: Hierarchical Evaluation of Segment Boundary Detection, in ISMIR , 2015.
- [6] M. Levy and M. Sandler: Structural segmentation of musical audio by constrained clustering, IEEE transactions on audio, speech, and language processing, 2008.

- [7] Carlos Hernandez-Olivan, Jose R. Beltran, David Diaz-Guerra: Music Boundary Detection using Convolutional Neural Networks: A comparative analysis of combined input features, *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 7, no 2, p. 78-88, 2021.
- [8] Brian McFee, D. Ellis: Analyzing Song Structure with Spectral Clustering, in *ISMIR* , 2014.
- [9] Meinard Muller, Nanzhu Jiang: A Scape Plot Representation for Visualizing Repetitive Structures of Music Recordings, *ISMIR* , 2012.
- [10] Morgan Buisson, Brian Mcfee, Slim Essid, Helene-Camille Crayencour: Learning Multi-Level Representations for Hierarchical Music Structure Analysis, *ISMIR* , 2022.
- [11] 林 亜紀 伊藤 貴之: Colorscore : MIDI を利用したクラシック楽曲構造の可視化と圧縮表示, *情報処理学会研究報告*, Vol.2010-MUS-86 No.27 , 2010.
- [12] Brian McFee, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenberg, Oriol Nieto: librosa: Audio and Music Signal Analysis in Python, *SciPy 14th Python in Science Conference*, 2015.
- [13] Alexandre Défossez: Hybrid Spectrogram and Waveform Source Separation: *ISMIR 2021 MDX Workshop*, p. 11, 2021.
- [14] Yu-Te Wu, Yin-Jyun Luo, Tsung-Ping Chen, I-Chieh Wei, Jui-Yang Hsu, Yi-Chin Chuang, Li Su: Omnizart: A General Toolbox for Automatic Music Transcription: *Artificial Intelligence (cs.AI); Audio and Speech Processing (eess.AS)*, 2021.

## 謝 辞

本研究を進めるにあたり，指導教員である北原鉄朗教授には丁寧かつ熱心なご指導を賜り，心より感謝申し上げます．また，共に研究を進めてきた同研究室の同期生および先輩方にも多大なる支援と助言をいただきました．深く感謝いたします．