

音楽基盤モデルにおける 音響特徴と内在音高螺旋の関係

八木 颯斗^{1,a)} 高道 慎之介^{1,2,b)} 佐藤りん³ 田中 啓太郎³ 森島 繁生³

概要: 我々の先行研究では、音楽基盤モデルの中間表現を解析することで、音高螺旋構造がモデル内部に内在的に形成されることを示した。しかし、その解析は単一の音色条件に限定されており、音色の違いが内在音高螺旋の明瞭さや形状にどのような影響を与えるかについては未解明であった。そこで本研究では、入力音の音色差および調波構造の違いに着目し、複数楽器の単音および調波成分を制御した人工音を用いて、それらが内在音高螺旋に与える影響を分析する。さらに、調波条件と螺旋らしさ指標である Helicality の対応を重回帰分析と統計的検定により解析し、内在音高螺旋の形成に寄与する調波成分を推定する。本手法は、音楽基盤モデルにおける音高と音色の関係性の理解を深めるとともに、音楽的概念に基づく内部表現解析に新たな視点を与える。

1. はじめに

近年、自己教師あり学習を用いた音楽基盤モデル (music foundation models) が目覚ましい発展を遂げている [1], [2], [3], [4], [5]。これらのモデルは、膨大な楽曲データから音楽の特徴を学習しており、生成だけでなく楽曲分析や検索といった広範なタスクにおいて高い性能を示している [5], [6]。一方で、これらのモデルが内部でどのような音楽的概念をどのように表現しているかについては、依然として十分に解明されていない。

我々は先行研究 [7] において、音楽基盤モデルの中間表現を解析することで、モデル内部に人間の音高知覚と類似した音高螺旋構造 [8] が出現することを示した。具体的には、音高のみが異なる音楽信号を入力とした際の中間表現を低次元化し、螺旋モデルをフィッティングすることで、螺旋構造らしさを定量的に評価する指標 (Helicality) を導入した。その結果、音高のような抽象的な概念はモデルの深い層においてより明確に表現されること、およびモデル

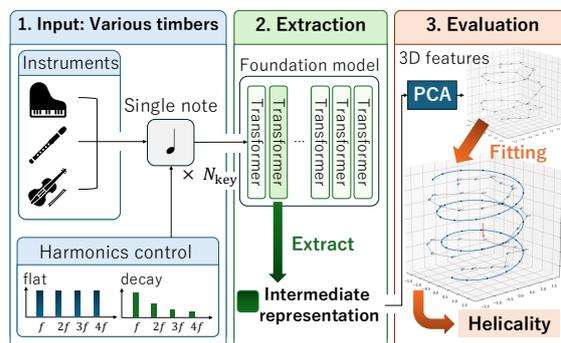


図 1: 各音色に対する内在音高螺旋の解析の流れ

サイズや階層構造が螺旋表現の明瞭さに影響を与えることが明らかとなった。以降では、この螺旋のことを**内在音高螺旋**と称する。

一方で、同研究 [7] では acoustic grand piano 合成音のみを対象としており、音色の違いが内在音高螺旋に与える影響については十分に検討されていない。一般に、音楽生成の制御を目的とした音高・音色の分離研究 [9] では、音高と音色は独立した属性として扱われることが多いが、実際には同一楽器であっても音高が変化することでスペクトル構造が変わり、知覚される音色も変化することが知られている [10], [11]。

そこで本研究では、複数の楽器および音色を対象として、内在音高螺旋の出現を系統的に分析する。具体的には、異なる音色の単音を入力とした際に得

¹ 慶應義塾大学
Keio University

² 東京大学
University of Tokyo

³ 早稲田大学
Waseda University

a) hayatobuti523@keio.jp

b) shinnosuke_takamichi@keio.jp

られる中間表現に対して螺旋モデルをフィッティングし、その形状パラメータおよび Helicality スコアを比較することで、音色の違いが内在音高螺旋に与える影響を明らかにする (図 1)。さらに、音色差の背後にある音響的要因を同定するため、基音に対して整数倍音やサブハーモニクスを付与した人工音を用い、調波条件と Helicality スコアの対応関係を重回帰分析および統計的検定により解析することで、内在音高螺旋の形成に寄与する調波成分を推定する。

本研究の目的は、音楽基盤モデルが音高情報をどのような条件のもとで螺旋構造として獲得するのかを明らかにするとともに、その形成に寄与する音響的要因を分析することである。本研究は、音楽基盤モデルにおける音高と音色の関係性の理解を深めるだけでなく、音楽的概念に基づく内部表現解析や、将来的な音楽生成モデルの制御性向上にも寄与することが期待される。

2. 関連研究

2.1 楽器の音色

楽器音は、基音と複数の倍音 (部分音) から構成される複合音であり、これらの周波数分布や強度分布、および時間的変化が音色を規定する主要な要因となる。倍音構造は、楽器の発音原理や振動体の性質に強く依存しており、弦楽器、管楽器、打弦楽器などの間で大きく異なることが知られている [10]。

弦楽器では、弦の振動モードにより理想的には整数次の倍音が生じるが、弦の剛性や減衰特性の影響により、特にピアノでは倍音が整数倍からわずかにずれる非調和性が現れることが知られている。擦弦楽器では、弓と弦の間の粘着・滑り運動により、鋸歯状波に近い振動が生じ、豊かな調和倍音を含むスペクトルが形成される。管楽器においては管の形状が倍音構造を決定し、円筒管では奇数次倍音が、円錐管では整数次倍音が顕著となる。さらに、電子楽器や音響合成音では、加算合成や周波数変調合成などにより倍音構造を意図的に制御することが可能であり、実在楽器の模倣だけでなく、現実には存在しないハイブリッドな音色を合成することもできる。

同一楽器であっても、演奏される音高によって倍音構造は変化する。共鳴特性や放射条件の違いにより、音高に応じて倍音の数や強度分布、スペクトル包絡が変化することが知られている。さらに、演奏強度の増加に伴い非線形性が強まり、高次倍音成分が相対的に強くなる。

図 2 に示すように、同一の音高 (C4) であっても、楽器によって基音近傍のエネルギー分布や倍音の立

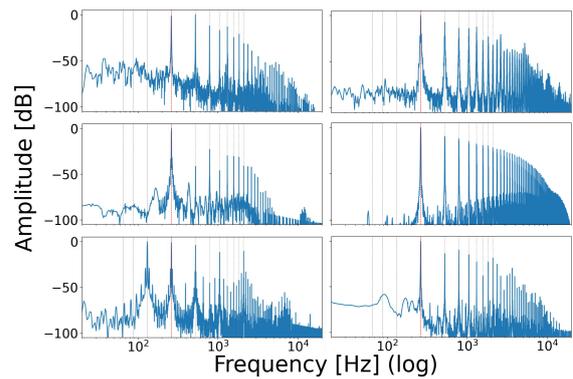


図 2: SynTheory [12] データセットの各楽器における C4 単音の周波数振幅スペクトル。赤の波線は基音を表し、黒の波線は基音に対する調波成分の位置を示す (サブハーモニクス: f_0/k ($k = 2, 3, 4$), 整数倍音: kf_0 ($k = 2, \dots, 8$)). (上段) 左: acoustic grand piano, 右: violin, (中段) 左: recorder, 右: sawtooth wave lead, (下段) 左: hammond organ, 右: acoustic steel guitar.

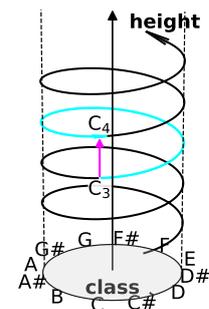


図 3: Shepard の音高螺旋構造 [8]。

ち上がり方が大きく異なる。このような楽器固有かつ音高依存的なスペクトル構造の違いは、音楽基盤モデルが獲得する音高表現の内部構造にも影響を与える可能性が高い。

2.2 内在螺旋構造

音高は、その高さ方向の連続性を表す pitch height と、オクターブ周期性に基づく循環構造を表す pitch class からなる二次元的な知覚構造を持つことが知られており、これらを統合した表現として音高螺旋 (pitch helix) が提案されている [8] (図 3)。近年では、このような音高の幾何学的構造が、ラベルなし音響データから教師なしに抽出可能であることも報告されている [13], [14]。

これに対し、我々の先行研究 [7] では、音楽信号そのものではなく音楽基盤モデルの中間表現に着目し、モデル内部に音高螺旋構造が内在的に形成されているかを解析した。具体的には、学習済み音楽基盤モデルに対して N_{key} 個の音高のみが異なる単音を入力

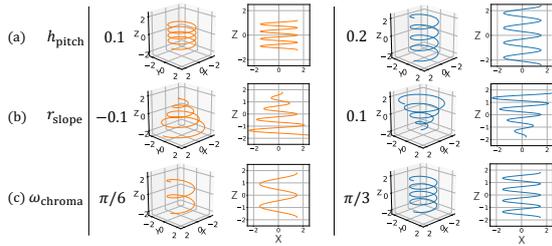


図 4: 式 (1) における各パラメータが螺旋形状に与える影響. 各行で一つのパラメータのみを変化させる: (a) h_{pitch} は高さ方向の間隔 (縦方向スケール) を制御する, (b) r_{slope} は円錐性を制御する (正: 半径が増加, 負: 半径が減少), (c) ω_{chroma} は 1 回転あたりの音高数 (角周波数) を制御する. 各列は, 3次元表示 (左) とその xz 平面への射影 (右) を示す. 比較の際, その他のパラメータは固定した.

し, 各層から抽出した中間表現を主成分分析 (PCA) により低次元化した上で, 3次元空間上にパラメトリックな音高螺旋モデルをフィッティングした.

音高インデックス t に対する螺旋モデル $\mathbf{y}(t) \in \mathbb{R}^3$ は, 高さ変化係数 h_{pitch} , 初期半径 r_0 , 半径変化係数 r_{slope} および角周波数 ω_{chroma} など複数のパラメータを用いて, 以下のように定式化される:

$$\mathbf{y}(t) = h(t) \cdot \mathbf{c} + r(t) (\cos \theta(t) \cdot \mathbf{u} + \sin \theta(t) \cdot \mathbf{v})$$

$$\text{where } \begin{cases} h(t) = h_{\text{pitch}} \cdot t + h_0 \\ r(t) = r_{\text{slope}} \cdot t + r_0 \\ \theta(t) = \omega_{\text{chroma}} \cdot (t - t_0) \end{cases} \quad (1)$$

ここで $\mathbf{c}, \mathbf{u}, \mathbf{v}$ は \mathbb{R}^3 の正規直交基底であり, それぞれ螺旋の中心軸および回転平面を定義する. 代表的なパラメータが螺旋形状に与える影響を図 4 に示す.

さらに, PCA で得た特徴量群と螺旋モデルの適合度を定量化する指標として, 先行研究 [14] に基づく Helicity スコア を用い, 3次元特徴量 $\{\mathbf{x}_t\}_{t=1}^{N_{\text{key}}}$ に対して以下のスコアを定義した:

$$\text{Helicity} = \left(\frac{1}{N_{\text{key}}} \sum_{t=1}^{N_{\text{key}}} |\mathbf{x}_t - \mathbf{y}(t)|^2 \right)^{-1} \quad (2)$$

この指標を用いることで, 層ごとおよびモデルごとに内在音高螺旋の明瞭さを評価した.

その結果, Jukebox [1] や MusicGen [3] といった音楽基盤モデルにおいて, 音高の螺旋構造が主として深い層で明瞭に出現することが確認された. また, モデルサイズや階層的アーキテクチャの違いによって, 螺旋構造の局在性や明瞭さが変化することが示され, 音高のような抽象的音楽概念がモデル内部で段階的に形成される可能性が示唆された.

一方で, 先行研究 [7] では acoustic grand piano の単音合成音のみを対象としており, 音色の違いや楽器間のスペクトル構造の差異が, 音高螺旋構造の形状や安定性にどのような影響を与えるかについては検討されていない. また, 螺旋構造の有無や明瞭さは評価されたものの, 螺旋モデルの形状を規定するパラメータ自体が持つ意味や, それらに基づく楽器間の関係性については未解明である.

そこで本研究では, 複数の楽器および音色を対象として音高螺旋構造の出現を再検証し, Helicity スコアに加えて螺旋モデルの形状パラメータの違いに着目することで, 音色や楽器の違いが音楽基盤モデル内部の音高表現に与える影響を体系的に分析する.

3. 分析手法

本研究では, 音楽基盤モデルの内部表現において, 音色や楽器の違いが Helicity スコアおよび音高螺旋構造の形状に与える影響を分析する. 以下に, 本研究で用いる手法の詳細を述べる.

3.1 種々の楽器への適用

本研究では, 図 1 に示すように, 先行研究 [7] で用いられた内在音高螺旋解析の枠組み (2.2 節) を, 複数の音色に対して同一の手法で適用する.

具体的には, 各入力音楽信号から得られる時系列の中間表現を時間方向に平均化し, 入力全体を代表する中間表現ベクトルとして扱う. この操作を音色条件ごとに適用することで, 各音色に対応する中間表現集合を得る. そして, 得られた中間表現集合に対しては, 各音色・各層ごとに PCA を適用する. 得られた特徴量は, 層および音色間での比較のために標準化した上で, 音高螺旋解析および Helicity スコア による評価を行う. 螺旋モデルおよび評価指標の詳細は先行研究 [7] を参照されたい.

以上の手順により得られた Helicity スコアを音色間で比較することで, 音色の違いが音楽基盤モデル内部に形成される内在音高螺旋の明瞭さに与える影響を分析する.

3.2 調波成分の寄与解析

本節では, 内在音高螺旋の明瞭さ (Helicity) が, 入力信号に含まれるどの調波成分によって強く駆動されるかを明らかにするための分析手法を述べる. 具体的には, 基音に対して特定の調波正弦波を加算した人工音を多数合成し, それぞれに対して音楽基盤モデルの中間表現から Helicity スコアを算出する. 得られた Helicity スコアと調波条件の対応関

係を、重回帰分析および統計検定により解析することで、螺旋スコアへ寄与する調波成分を同定する。

3.2.1 調波構造を制御した人工音の合成

基本周波数を f [Hz] とし、時間 t における人工音信号 $s_{\mathcal{H},f}^{(\text{tilt})}(t)$ を以下で定義する：

$$s_{\mathcal{H},f}^{(\text{tilt})}(t) = \sum_{h \in \mathcal{H}} a_h^{(\text{tilt})} \sin(2\pi fht) \quad (3)$$

ここで \mathcal{H} は付与する調波インデックスの集合である。本研究では基音成分 $h = 1$ は常に含まれるものとし、整数倍音 $h = 2, 3, \dots$ およびサブハーモニクス $h = \frac{1}{2}, \frac{1}{3}, \dots$ の有無を条件として \mathcal{H} を変化させる。採りうる h の全体集合を \mathcal{H}_{all} とすると \mathcal{H} はその部分集合 $\mathcal{H} \subset \mathcal{H}_{\text{all}}$ となる。調波インデックス集合 \mathcal{H} の違いが周波数振幅スペクトルに与える影響を図 5 に示す。

tilt は調波強度の傾き (spectral tilt) であり、spectral tilt 条件の集合を $\mathcal{T} = \{\text{flat}, \text{decay}\}$ とする。調波強度の傾きを制御するため、各調波の振幅係数 $a_h^{(\text{tilt})}$ として以下の 2 条件を用いる：

$$a_h^{(\text{flat})} = 1 \quad (4)$$

$$a_h^{(\text{decay})} = \min(h, 1/h) \quad (5)$$

式 (4) は全調波を等強度で加算する条件であり、式 (5) は、基音から周波数比が離れるほど (高次側・低次側の双方で) 振幅が減衰するように定義した対称な tilt 条件である (図 6)。同一の調波インデックス h であっても、flat 条件と decay 条件は異なる音色的特性を与えるため、本分析ではこれらを別個の説明変数として扱う。

3.2.2 調波条件ごとの Helicity スコア算出

調波条件 (\mathcal{H}, tilt) に基づいて、3.1 節と同様に、あらかじめ定めた音高範囲に含まれる N_{key} 個の異なる音高の音を合成する。この人工音の集合を $\mathcal{S}_{\mathcal{H}}^{(\text{tilt})} = \{s_{\mathcal{H},f}^{(\text{tilt})}\}_{f \in \mathcal{F}}$ とする。 \mathcal{F} は、定めた音高範囲に含まれる、各音高の基本周波数の集合である。各人工音を音楽基盤モデルに入力して、音楽基盤モデルの各 Transformer 層から中間表現を抽出し、Helicity スコアを算出する。

本節の統計解析では、人工音の集合 $\mathcal{S}_{\mathcal{H}}^{(\text{tilt})}$ ごとの層分布を要約した以下の 2 種類の従属変数を用いる。Transformer 層 ℓ における Helicity スコアを $H_{\ell}(\mathcal{S}_{\mathcal{H}}^{(\text{tilt})})$ とすると、

$$y_{\text{sum}}(\mathcal{S}_{\mathcal{H}}^{(\text{tilt})}) = \sum_{\ell \in \mathcal{L}} H_{\ell}(\mathcal{S}_{\mathcal{H}}^{(\text{tilt})}) \quad (6)$$

$$y_{\text{peak}}(\mathcal{S}_{\mathcal{H}}^{(\text{tilt})}) = \max_{\ell \in \mathcal{L}} H_{\ell}(\mathcal{S}_{\mathcal{H}}^{(\text{tilt})}) \quad (7)$$

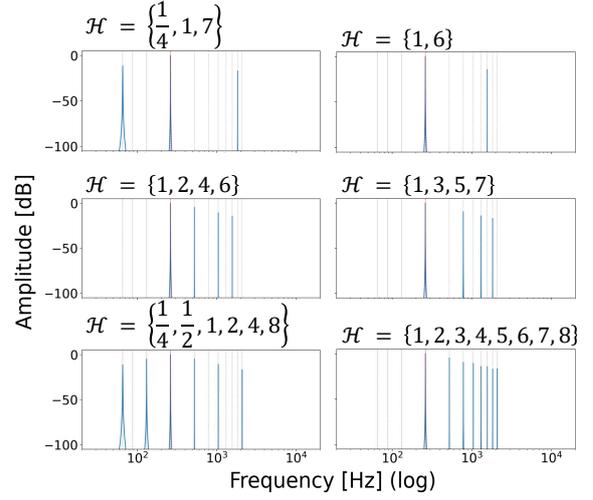


図 5: 調波インデックス集合 \mathcal{H} の違いによって生じる周波数振幅スペクトルの例。基本周波数は C4 に固定されており、spectral tilt 条件は decay に統一されている。

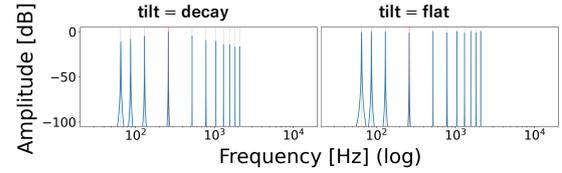


図 6: spectral tilt 条件の違いによって生じる周波数振幅スペクトルの例。基本周波数は C4 に固定されており、両条件で用いた調波インデックス集合は $\mathcal{H} = \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1, 2, 3, 4, 5, 6, 7, 8\}$ である。

をそれぞれ「螺旋らしさの総量 (層合計 Helicity)」および「最尤の螺旋らしさ (層最大 Helicity スコア)」と定義する。ここで \mathcal{L} は解析対象とする Transformer 層の集合である。 y_{sum} は層全体にわたる内在音高螺旋の総合的な強さを表し、 y_{peak} は最も明瞭な内在音高螺旋が出現する層における強さを表す。

3.2.3 重回帰分析と統計的検定

調波条件 (\mathcal{H}, tilt) を、各調波条件と spectral tilt 条件の組に対応するバイナリ変数により表現し、Helicity スコアとの関係を重回帰分析によりモデル化する。まず、バイナリ変数を以下のように定義する：

$$x_{h,\text{ti}} = \begin{cases} 1 & (h \in \mathcal{H} \wedge \text{ti} = \text{tilt}) \\ 0 & \text{Otherwise} \end{cases} \quad (8)$$

例えば $(h, \text{ti}) = (2, \text{flat})$ のとき、すなわち $x_{2,\text{flat}}$ は、第 2 倍音成分が含まれ、かつ spectral tilt 条件が flat である調波成分の有無を表す説明変数である。そして、従属変数 $y \in \{y_{\text{sum}}, y_{\text{peak}}\}$ に対し、以下の線形モデルを仮定する：

$$y = \beta_0 + \sum_{(h,ti) \in (\mathcal{H}_{\text{all}} \setminus \{1\}) \times \mathcal{T}} \beta_{h,ti} x_{h,ti} + \varepsilon \quad (9)$$

ここで β_0 は切片, $\beta_{h,ti}$ は調波 h を tilt 条件 ti で付与したときの寄与を表す回帰係数, \times は直積集合の演算子, ε は誤差項である. なお, 回帰から基音 $h = 1$ の条件を除くため, $\mathcal{H}_{\text{all}} \setminus \{1\}$ について総和している. 種々の調波条件に対し $(x_{h,ti}, y)$ を求め, 係数 $\beta_{h,ti}$ を最小二乗法により推定する. その後, モデル全体の有意性を評価するために F 検定を行い, さらに各係数について t 検定を行うことで, 対応する (h, ti) が螺旋スコアに有意な影響を与えているかを評価する.

以上により, (i) 層全体にわたる螺旋構造を強める調波成分 (y_{sum}), および (ii) 特定層における螺旋構造のピーク形成に寄与する調波成分 (y_{peak}) を, 調波成分と spectral tilt の組として分離して同定することが可能となる.

4. 実験的評価

4.1 実験条件

本研究では, 先行研究 [7] と同一の解析手順を基本としつつ, 入力データを拡張することで, 音色および調波構造の違いが内在音高螺旋に与える影響を検証した. 以下に, 本研究で用いた実験条件を示す.

4.1.1 モデル条件

音楽基盤モデルとして, 学習済みの Jukebox [1] top-level decoder (5B) を用いた. モデルの入手方法と中間表現の抽出については, 先行研究 [7] と同じである.

4.1.2 データ条件

本研究では, 音高螺旋構造の解析のために, 以下の 2 種類のデータセットを用いた. 両データセットにおいて, 音楽条件, およびオーディオ条件, PCA および射影条件については, 共通の条件を採用し, それぞれの条件は先行研究 [7] と同じである.

楽器単音データセット: 合成音楽理論データセット SynTheory [12] の Notes データセットに含まれる, 92 種類の楽器を解析対象とした. 各楽器は General MIDI に基づく音色プログラム番号および楽器カテゴリを有する.

人工音データセット: 式 (3) で定義した人工音の集合. 以下に示すように, 整数倍音 (最大 8 倍音) およびサブハーモニクス ($h = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}$) を含む複数の調波インデックス集合 \mathcal{H} を設定した.

- 全倍音を含む
- 奇数倍音のみを含む
- 偶数倍音のみを含む

- 2 の冪倍音のみを含む
- 特定の 1 つの倍音のみを含む
- 特定の 2 つの倍音のみを含む
- サブハーモニクスを含む

各 \mathcal{H} に対し spectral tilt 条件 $\mathcal{T} = \{\text{flat}, \text{decay}\}$ を適用し, 最終的に 137 種類の人工音集合を作成した.

4.1.3 フィッティング条件

式 (1) のパラメータの最適化手法は, 先行研究 [7] と同一である. ただし, 本研究ではデータセットごとに試行回数を調整し, 楽器単音データセットに対しては, 試行回数 1000 回および 2000 回のパラメータ最適化をそれぞれ設定し, 各設定について異なる乱数シードを用いて 3 回ずつ最適化を行った. なお, 試行回数 1000 回および 2000 回の場合で, パラメータ探索範囲はそれぞれ異なる設定を用いた. 一方, 人工音データセットに対しては, 試行回数 1000 回のパラメータ最適化を, 異なる乱数シードを用いて 3 回繰り返して行った. いずれの条件においても, 各最適化試行で得られた Helicality スコアのうち, 最も高い値を記録したものを当該条件のスコアとして採用した. パラメータ探索範囲の詳細は付録 A.1 を参照されたい.

4.2 調査内容

実験では, 以下の内容を調査する.

楽器単音データセットを用いた評価. 各楽器の音色が内在音高螺旋にどう影響するかを調査する.

- **層最大 Helicality スコアは楽器ごとにどう異なるか? (4.3.1 節):** 楽器ごとのスコアを比較して, 音色と内在音高螺旋の関係を明らかにする.
- **Helicality スコアは各層にどう分布するか? (4.3.2 節):** 層最大だけでなく全ての層のスコアを分析し, 内在音高螺旋がどの層で顕在化しやすいかを明らかにする.
- **ピーク層と螺旋パラメータはどう関係するか? (4.3.3 節):** ピーク層における螺旋パラメータを比較し, 層位置や楽器条件とパラメータの対応を明らかにする.

人工音データセットを用いた評価. 楽器音を構成する調波構造と spectral tilt が, 内在音高螺旋にどう影響するかを調査する.

- **どの調波成分あるいは強度が内在音高螺旋を構成するか? (4.4.1 節):** 調波成分の有無および強度を説明変数とする重回帰分析により, 層全体の Helicality 総和スコアと層最大 Helicality スコアに対して, 正/負に寄与する調波成分を明らかにする.

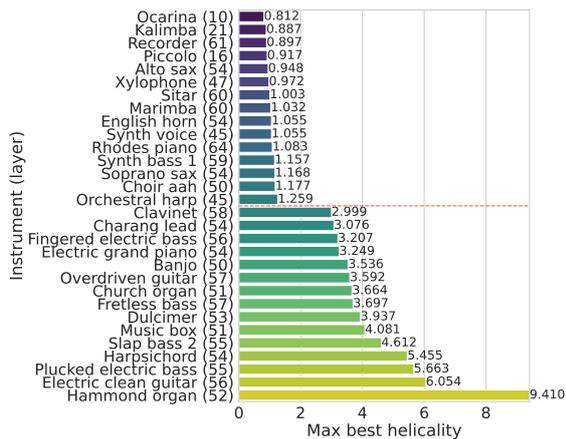


図 7: 層最大 Helicity スコア上位 15 楽器および下位 15 楽器. 赤の波線はそれぞれの境界を表す. 括弧内の整数値は, 最大スコアをとった Transformer 層のインデックスを表す.

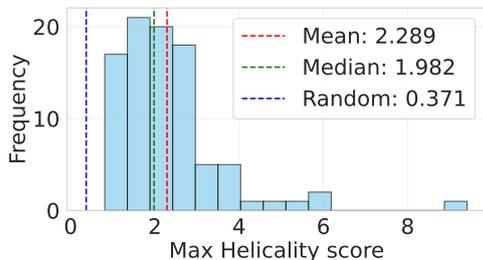


図 8: 全楽器における層最大 Helicity スコアの分布. 破線は平均値, 中央値, および 3 次元乱数に対する層最大 Helicity スコアの基準値を示す.

- **調波構造は Helicity スコア分布をどのように変えるか? (4.4.2 節):** 各層に対する内在音高螺旋の分布を調査し, 螺旋構造が浅い層から顕在化する調波構造条件を明らかにする.
- **調波周波数は螺旋の周期をどう規定するか? (4.4.3 節):** Helicity スコア分布のピーク層で推定される角周波数 ω_{chroma} の分布を調波構造ごとに比較し, 支配的な周期の選択が調波構造とどのように関係するかを明らかにする.

4.3 楽器に関する分析

4.3.1 楽器別の層最大 Helicity スコア

音色と内在音高螺旋の関係を明らかにするために, 楽器単音データセットに含まれる各楽器の層最大 Helicity スコアを計算した.

図 7 に, 層最大 Helicity スコア上位 15 楽器および下位 15 楽器, 図 8 に全楽器のスコアの分布を示す. Helicity スコアは楽器間で大きなばらつきを示しており, 最小値は ocarina における 0.812, 最大値は hammond organ における 9.410 である. 全

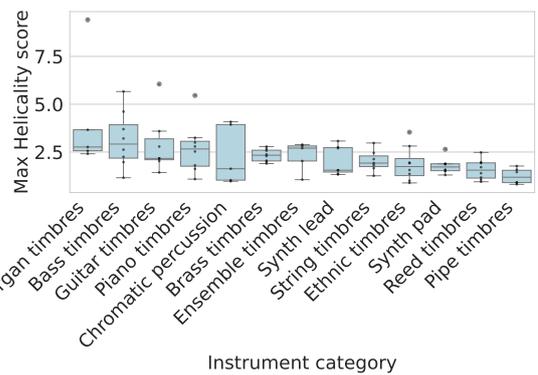


図 9: 楽器カテゴリごとの層最大 Helicity スコアの分布. 点は各楽器のスコアを表す.

体の中央値は 1.970, 平均は 2.285 であり, 多くの楽器のスコアは 1 から 3 の範囲に分布している. 一方で, 一部の楽器が極めて高いスコアを示しており, 分布は右に裾の長い形状を持つ. また, 3 次元乱数に対するスコア 0.371 (詳細は付録 A.2) と比較すると, 全ての楽器はこれを上回っている. 以上より, 内在螺旋構造は, 特定楽器 [7] のみで表れるものではなく, 多くの楽器において現れることを示している.

楽器ごとのスコアの違いを分析するため, General MIDI に基づく楽器カテゴリごとにスコアを集約した結果を図 9 および表 1 に示す. Organ timbres は平均が最も高く分散も大きいことから, カテゴリ内のばらつきが顕著である. Bass timbres や guitar timbres, piano timbres は中程度の平均だが大きな分散であった. 一方, brass timbres や string timbres では分散が小さく, カテゴリ内で比較的安定したスコア分布を示した. これより, 内在音高螺旋の形成には, 楽器カテゴリに基づく音色特性に対応した大域的な傾向と, 同一カテゴリ内における個別楽器固有の音響的要因が, それぞれ寄与している可能性が示唆される.

4.3.2 層別の Helicity スコア

4.3.1 節では層最大 Helicity スコアのみを議論した. ここでは, 最大スコアではなく各 Transformer 層の Helicity スコアを計算し, どの層で内在螺旋構造が顕在するのかを分析した.

楽器全体に対する各層の傾向. まず, 全体的な傾向を確認するため, 全楽器における層別 Helicity スコアの分布を計算した (図 11). スコアの層別平均は層が深くなるとともに緩やかに上昇し, 後半の層において顕著に増大する. また, 第 20 層から第 30 層付近にも小さな局所の上昇が観測される. この結果より, スコアの局所的なピークが層前半と後半にそれぞれ 1 箇所ずつ存在することを仮定して, 以

表 1: 楽器カテゴリごとの最大 Helicity に関する統計量.

カテゴリ	平均値	標準偏差	最小値	最大値	楽器数
Organ timbres	4.16	2.97	2.41	9.41	5
Bass timbres	3.14	1.48	1.16	5.66	8
Guitar timbres	2.89	1.55	1.43	6.05	7
Piano timbres	2.69	1.34	1.08	5.46	8
Chromatic percussion	2.33	1.56	0.97	4.08	5
Brass timbres	2.32	0.34	1.90	2.79	8
Ensemble timbres	2.30	0.77	1.06	2.89	5
Synth lead	2.04	0.76	1.33	3.08	7
String timbres	2.00	0.54	1.26	2.97	7
Ethnic timbres	1.88	0.91	0.89	3.54	8
Synth pad	1.77	0.41	1.30	2.64	8
Reed timbres	1.58	0.53	0.95	2.48	8
Pipe timbres	1.24	0.41	0.81	1.77	6
All instruments	2.29	1.27	0.81	9.41	92

降の議論を進める. この議論は, 内在音高螺旋が明瞭な螺旋明瞭群 45 楽器のみを対象とする. この選定基準は, 図 8 に示す楽器別の層最大 Helicity スコアが 2.0 以上とした.

前半層・後半層の比較. Transformer 全 72 層を第 36 層を境界として前半層と後半層に二分し, 半層それぞれで最大の Helicity スコアを計算した. 半層それぞれの最大スコアを比較し, スコアの大きい Transformer 層を最大ピーク層 (max peak), もう一つをサブピーク層 (sub peak) と定義した. 図 10 に, 螺旋明瞭群に含まれる各楽器の最大ピーク層とサブピーク層の分布を示す. 先行研究 [7] でも示されたように, 内在音高螺旋は主として後半層において最も明瞭に形成される一方で, 少数だが前半層に形成する楽器も存在する. また後半層に明瞭に形成する楽器に着目すると, 図 10 の点の大きさが大きい, すなわち最大ピークとサブピークのスコア値の比が大きいことから, 前半層のやや不明瞭な螺旋と後半層の顕著に明瞭な螺旋から構成されることが伺える.

4.3.3 ピーク層とパラメータの関係

4.3.2 節で言及したピーク層においてどのような螺旋パラメータが推定されたのか, また, その推定値は何の要素によって決定されるのかを調査する.

ピーク層における角周波数. 図 12 に, 最大/サブピーク層インデックスと, その層において推定された角周波数 ω_{chroma} の散布図を示す. まず, 全ての楽器・条件において角周波数は $\pi/3$ または $\pi/6$ のいずれかに近い値を取ることが確認できる. これは, 内在音高螺旋が 1 回転/オクターブ ($\omega_{\text{chroma}} \approx \pi/6$) または 2 回転/オクターブ ($\omega_{\text{chroma}} \approx \pi/3$) という二種類の幾何構造に強く制約されていることを意味

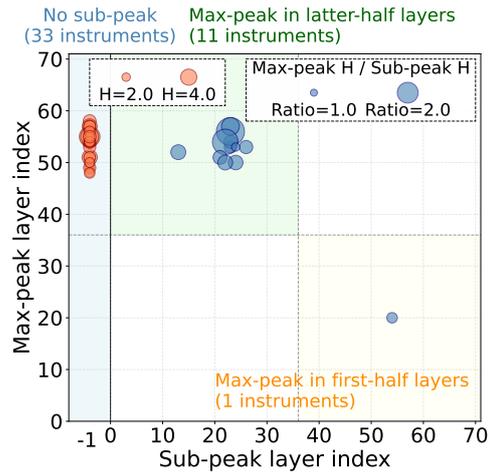


図 10: 螺旋明瞭群における Helicity スコア (H) のピーク層の散布図. 青色プロット大きさは, 最大ピーク c とサブピーク層のスコアの比を表している. 横軸の -1 における点は sub-peak がないことを表し, この点の大きさは 層最大 Helicity スコアを表している.

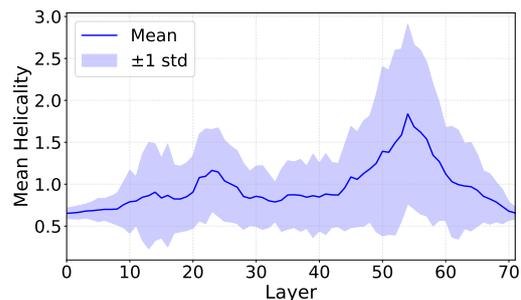


図 11: 全楽器における層別 Helicity スコアの平均.

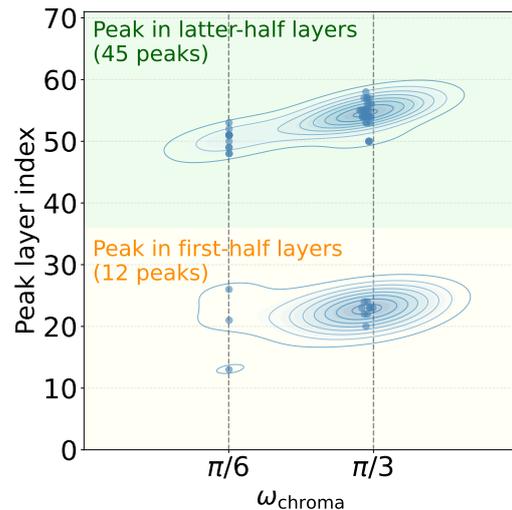


図 12: ピーク層番号と ω_{chroma} の関係.

する. 次に後半層インデックスとの関係を見ると, $\omega_{\text{chroma}} \approx \pi/3$ の楽器は, $\omega_{\text{chroma}} \approx \pi/6$ の楽器に比べてより深い層にピークを持つことがわかる.

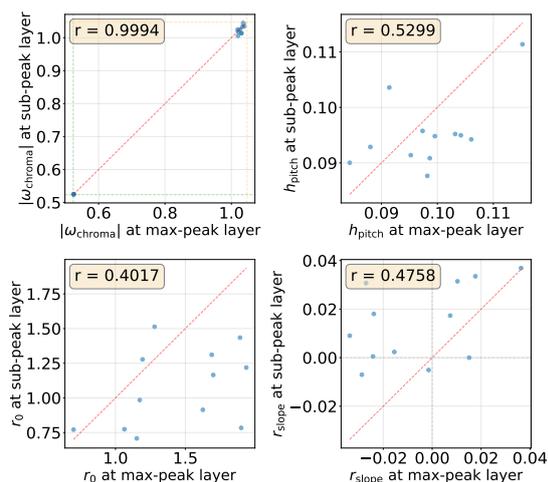


図 13: 最大ピーク層とサブピーク層を持つ楽器における、各螺旋パラメータの散布図。赤い破線は $y = x$ を示す。

最大ピーク層とサブピーク層の螺旋パラメータの比較. 最大ピーク層とサブピーク層を持つ楽器に対し、各ピーク層の螺旋パラメータを比較した結果を図 13 に示す。特筆すべき点として、 ω_{chroma} は極めて高い相関 ($r \approx 0.999$) を示し、同一楽器内で ω_{chroma} は変化しないことがわかる。故に、 ω_{chroma} は (ピーク) 層には依存せず、楽器固有の音響的特性に依存している可能性を示している。一方で他の螺旋パラメータについては、相関が中程度にとどまっており、パラメータが層に依存する可能性を示している。

ピーク層インデックスと螺旋パラメータの関係. ピーク層の螺旋パラメータについて、前半層と後半層に分けてその分布を調べた結果を図 14 に示す。この図より以下のことが明らかである。

- **半径変化係数 r_{slope} :** 前半層では非負値 (円筒型や、音高に対し半径が増加する円錐型) をとり、後半層では負値 (音高に対し半径が減少する円錐型) もとる。すなわち、深い層のピークでは、より多様な螺旋形状が形成されることを示唆する。
- **初期半径 r_0 :** 前半層は後半層に比べ小さい値をとる。これは、深い層ほど螺旋の回転平面上での中心軸からの広がりが大きく、pitch class 方向の分離がより大きい表現になっている可能性を示唆する。
- **高さ変化係数 h_{pitch} :** 前半層と後半層の間で顕著な分布差は観測されていない。このことから、 h_{pitch} はピーク層の位置に依存せず比較的安定したパラメータである可能性が示唆される。

この結果は、最大ピーク層の位置によって螺旋パラ

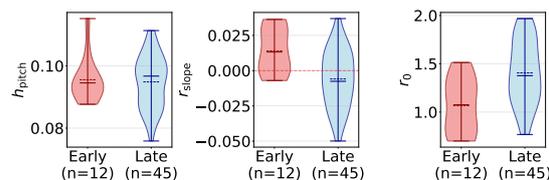


図 14: Early/late ピークにおける各パラメータ分布の比較。

表 2: 人工音における層最大 Helicity の上位例および下位例

順位	調波集合 \mathcal{H}	Spectral tilt peak layer	y_{peak}
1	$\{\frac{1}{2}, 1, 2, 4, 8\}$	decay	52 7.744
2	$\{\frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8\}$	decay	51 7.688
3	$\{\frac{1}{2}, 1, 2\}$	flat	50 6.906
4	$\{\frac{1}{2}, 1, 2\}$	decay	56 6.469
5	$\{\frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8\}$	flat	53 5.097
⋮	⋮	⋮	⋮
133	$\{1\}$ (純音)	—	21 0.619
134	$\{1, 6, 8\}$	flat	36 0.594
135	$\{1, 6, 8\}$	decay	1 0.593
136	$\{1, 7\}$	decay	0 0.589
137	$\{1, 7, 8\}$	decay	0 0.583

メータの範囲が異なることを示唆している。

4.4 調波構造の寄与分析

4.3 節では、楽器単音データセットを用いて音色が内在音高螺旋の形成に寄与するかを調査した。それを受けこれ以降は人工音データセットを用いて、楽音を成す調波構造と spectral tilt が内在音高螺旋の形成に寄与するかを調査する。

4.4.1 調波構造と Helicity スコアの関係

3.2 節で述べた重回帰分析に基づき、各調波成分の有無と spectral tilt が内在音高螺旋の明瞭さ寄与しているかを分析した。重回帰分析にあたり求めた各条件の層最大 Helicity スコアの例を表 2 に示す。F 検定の結果、層平均 Helicity スコア y_{sum} と層最大 Helicity スコア y_{peak} のいずれへの回帰についても統計的有意差が認められたため、各回帰係数に対する t 検定の結果に基づき、各条件の寄与を解釈する。図 15 と図 16 はそれぞれ、 y_{sum} と y_{peak} を従属変数とした場合の回帰係数と p 値である。これらの図に基づいて結果を考察する。なお、両方の図は似た結果を示しており、以降に述べる考察は両方の図に適用される。

オクターブ等価の調波が螺旋を形成する. 第 $\{2, 4, 8\}$ 倍音および $\{1/2, 1/4\}$ サブハーモニクス成分は、大きな正の係数を示しており、 p 値もほとんど

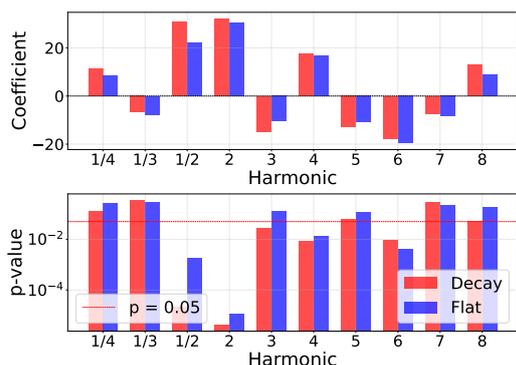


図 15: 総和 Helicity スコア y_{sum} に対する重回帰係数。下段は各係数に対応する t 検定の p 値を示しており、赤の波線は有意水準 $p = 0.05$ を表す。

の場合で 0.05 を下回っている。この結果は、人間の音高知覚におけるオクターブ等価性と整合しており、音楽基盤モデル内部においても、周波数比が 2 の冪乗で表される調波成分が、内在音高螺旋の形成において重要な役割を果たしている可能性を示唆している。この結論は楽器単音データベースを用いた結果とも整合しており、Helicity スコア (図 7) の最も高い hammond organ は、オクターブ等価の調波が支配的なスペクトル (図 2) を有する。

オクターブ等価でない調波は螺旋形成を阻害する。 対して、残る第 {3, 5, 6, 7} 倍音および 1/3 サブハーモニクス成分は総じて負の値を取る。特に、第 6 倍音は有意に負の寄与を示している。すなわち、オクターブ等価でない調波成分によって、螺旋の形成が阻害されていることを示唆する。Helicity スコア (図 7) の低い楽器である recorder では、確かに、オクターブ等価でない調波が支配的 (図 2) である。

周波数に対して減衰する調波は形成を補助する。 spectral tilt 条件の違いに着目すると、オクターブ等価の全調波において flat 条件よりも decay 条件が大きい係数を示している。全調波が同じ強度を持つ flat 条件に対し、楽音のように高周波数で減衰する decay 条件が優れることは、音楽基盤モデルが楽音らしい強度分布から螺旋を形成する可能性を示唆している。

調波を有さない信号は螺旋を形成しづらい。 表 2 に示すように、純音は低い Helicity スコアを示す。また、最も低い Helicity スコアを示した ocarina は、単一モードの Helmholtz 共鳴器として振る舞い [10]、明確な整数倍音系列を持たないことから、純音入力に近い条件として Helicity スコアが低くなると考えられる。すなわち、調波を有さない信号は螺旋を形成しづらいと言える。この結論は、前述し

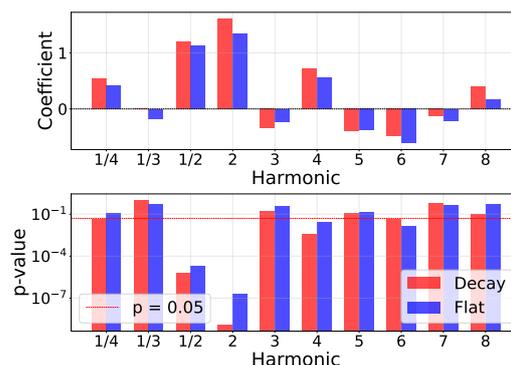


図 16: 層最大 Helicity スコア y_{peak} に対する重回帰係数。

た調波の有無とも整合する。

4.4.2 調波構造とピーク層の関係

調波構造の違いが Helicity スコアのピーク層の出現傾向にどのように影響するかを分析した。4.3.2 節にて仮定した「スコアの局所的なピークが層前半と後半にそれぞれ 1 箇所ずつ存在すること」を導入し、ピークの発生条件を人工音データセットを用いて調査した。137 種類の人工音集合のうち、螺旋明瞭群の定義に該当する 30 種類を対象とした。

楽器単音と同様に後半層に最大ピーク層が現れる。

図 10 と同様に前半層と後半層に分け、最大ピーク層とサブピーク層の分布を図 17 に示す。図 10 と図 17 を比較すると、最大ピーク層が後半層に現れやすいことは共通することがわかる。

オクターブ等価の調波がサブピーク層を生む。 サブピーク層を生じさせる調波条件を詳細に調査するために、サブピーク層を持つ条件を集計した結果を表 3 に示す。サブピーク層が生じるのは、人工音がオクターブ等価の調波のみで構成されているときのみである。すなわち、前半層における Helicity スコアの局所的な上昇は、オクターブ等価の調波構造に条件づけられた現象であることを示唆している。この傾向は、図 18 から確認できる。

オクターブ等価でない調波はサブピーク層を消す。

同表より、オクターブ等価でない調波を 1 つでも含むとサブピーク層は発生しないことがわかる。4.4.1 節において述べた「オクターブ等価でない調波は螺旋形成を阻害する」と対応付けると、サブピーク層が消失することと、最大ピーク層の螺旋らしさが減少することに何らかの関係があると考えられる。

4.4.3 調波構造と ω_{chroma} の関係

4.4.2 節と同様に、最大ピーク層における ω_{chroma} と調波構造の関係を分析した。

楽器単音と同様に特定の角周波数を持つ。図 19

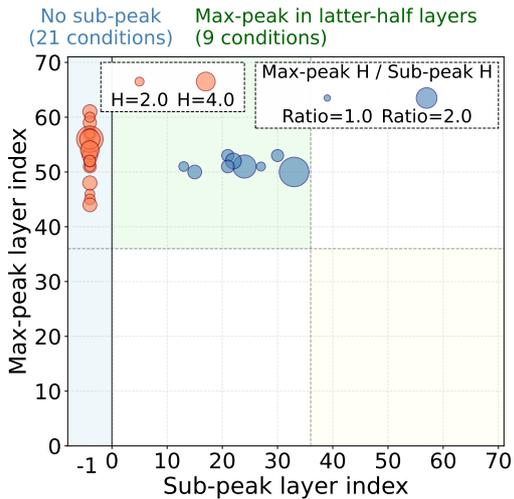


図 17: 螺旋明瞭群における人工音の Helicity スコアのピーク層の散布図. 横軸の -1 における点は sub peak がないことを表す.

表 3: サブピーク層の有無と調波構造の関係

調波構造の分類	該当数	サブピークあり内数
2 の冪倍音のみ		
第 2・第 4 倍音を両方含む	8	8
第 2 または第 4 倍音を含まない	9	1
非オクターブ関係の倍音を含む	13	0
合計	30	9

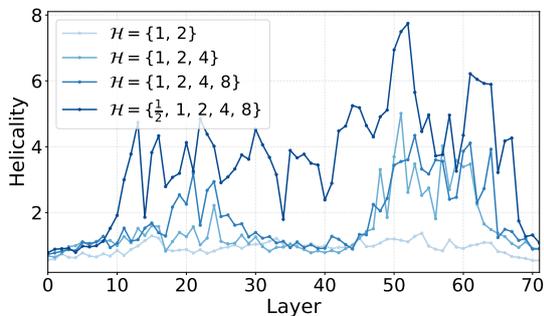


図 18: 2 の冪倍音の構成が異なる人工音に対する Helicity スコアの層方向変化. 調波インデックス集合 \mathcal{H} は, $\{1, 2\}$, $\{1, 2, 4\}$, $\{1, 2, 4, 8\}$, および $\{\frac{1}{2}, 1, 2, 4, 8\}$ を示す. 2 の冪倍音が増加するにつれて, 前半層における Helicity スコア の局所的な上昇が顕著になることが分かる.

表 4: 調波構造別に見たピーク層における ω_{chroma}

調波構造の分類	サンプル数	支配的周期	その割合
2 の冪倍音のみ	26	$\pi/6$	1.00
非 2 冪倍音を含む	13	$\pi/3$	0.77

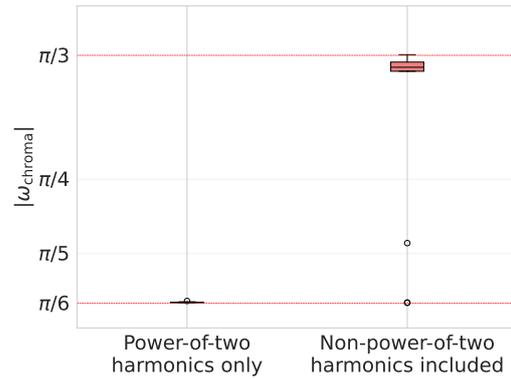


図 19: ピーク層における $|\omega_{\text{chroma}}|$ の分布. 赤の破線は $\pi/6$ および $\pi/3$ を示す.

は, オクターブ等価の調波のみで構成される人工音とそれ以外の人工音に分けて $|\omega_{\text{chroma}}|$ の分布を示している. この図から ω_{chroma} は主に $\pi/6$ または $\pi/3$ のいずれかに集中することが確認できる. この傾向は, 図 12 に示した楽器単音の結果と同様である.

オクターブ等価の調波がオクターブ等価の螺旋を形成する. 図 19 において, オクターブ等価の調波のみで構成される人工音では, 例外なく $|\omega_{\text{chroma}}| \approx \pi/6$ (1 回転/オクターブ) が観測される. $\pi/6$ は Shepard の pitch helix [8] に対応することを鑑みれば, オクターブ等価の調波がオクターブ等価を持つ螺旋を構成すると言える.

オクターブ等価でない調波がオクターブ等価でない螺旋を形成する. オクターブ等価でない調波を含む人工音では, 一部の例外を除き $|\omega_{\text{chroma}}| \approx \pi/3$ が主に観測される. $|\omega_{\text{chroma}}| = \pi/3$ は, 2 回転/オクターブの螺旋構造に対応し, トライトーン関係の音が螺旋上で同一角度に配置される構造を意味する. この例外を列挙したものが表 5 である. $|\omega_{\text{chroma}}| \approx \pi/6$ となる例外はオクターブ等価の調波が支配的であるため, オクターブ等価の調波のみの人工音と同じ結果になったと考えられる. $|\omega_{\text{chroma}}| \approx \pi/5$ の例外は, 3 倍音 (完全五度上) と $1/3$ 次倍音 (完全四度下) を同時に含む極めて特殊な構成である. この例について, ω_{chroma} を $\pi/6$ や $\pi/3$ に固定して螺旋パラメータを再推定したところ Helicity スコア が顕著に劣化したことから, $\pi/5$ に近い構造が実際に形成されていると考えられる.

以上の結果から, ω_{chroma} は入力音の調波構造, 特にオクターブ等価の調波構造に強く寄与される可能性が高いと考えられる. この知見は, 音楽基盤モデルにおける音高表現が, 単なる周波数スケールの違いとしてではなく, 入力音の調波構造が内包する音程関係に応じて, 異なる周期性を持つ音高螺旋とい

表 5: 非 2 の冪倍音を含む音における ω_{chroma} の例外

調波構造	tilt	$ \omega_{\text{chroma}} $	Helicity
基音 + 2, 7 倍音	decay	$\approx \pi/6$	2.174
基音 + 2, 4, 6, 8 倍音	decay	$\approx \pi/6$	2.625
基音 + 3, 1/3 倍音	decay	$\approx \pi/5$	2.684

う幾何構造として表現される可能性を示唆している。

5. まとめ

本研究では、音楽基盤モデルの内部表現における内在音高螺旋の形成メカニズムを明らかにすることを目的として、多様な楽器音および調波構造を制御した人工音を用いた網羅的な解析を行った。

まず、実楽器音を対象とした分析により、内在音高螺旋の明瞭さを表す Helicity スコアは楽器音色によって大きく変動することを確認した。この結果は、内在音高螺旋が特定の音色条件に限定された現象ではなく、多様な楽器音において形成されう一方、その明瞭さが音響的特性に依存することを示している。

次に、調波構造を制御した人工音を用いた重回帰分析を通じて、Helicity スコアの差異を生む音響的要因を定量的に評価した。その結果、オクターブ等価に対応する 2 の冪倍音成分が内在音高螺旋の形成に対して有意に正の寄与を示す一方、非オクターブ関係にある倍音成分は負の寄与を示すことが明らかとなった。

さらに、層方向および螺旋形状パラメータの解析から、先行研究 [7] と同様に、内在音高螺旋は主として Transformer の深い層において最も明瞭に形成されることを確認した。加えて、一部の条件では浅い層と深い層の双方にピークを持つ傾向が観測された。また、螺旋の回転角 ω_{chroma} が入力音の調波構成に応じて変化することから、モデルが入力音楽信号の音響的特徴に適応して、内部の音高表現を幾何学的に表現している可能性が示唆された。

一方で、実楽器音における結果を見ると、electric clean guitar など、Helicity スコアが高い楽器の中には、必ずしも 2 の冪倍音成分が支配的とは言えない例が存在する。このことは、調波構造が内在音高螺旋の形成において重要な役割を果たす一方で、それのみが十分条件ではないことを示している。

今後は、人工音の解析で得られた知見を踏まえ、実楽器音においても倍音分布やオクターブ等価性指標などを定量化し、それらと Helicity スコア や螺旋形状パラメータとの関係を体系的に分析する必要がある。加えて、時間的変動の大きさやアタック特性

など、調波構造以外の音響的要因が内在音高螺旋の形成に与える影響についても、今後詳細に検討することが重要であると考えられる。

本研究の知見は、音楽基盤モデルが音高を単なる周波数スケールの違いとしてではなく、調波構造に裏付けられた幾何学的構造として内部表現していることを示すものである。さらに、文献 [15] は、人間の音高に関する知覚特性が、音色への親和性よりも倍音の数や強度分布といった音響的構造と密接に関連していることを示している。この知見を踏まえると、本結果は、2 の冪倍音が支配的な音色において、人間の音高知覚に内在するオクターブ等価性や螺旋構造が、より明瞭に知覚・反映されう可能性を示唆するものであり、音楽基盤モデルの内部解析を通じて、人間の音高知覚特性に対する新たな理解を与える可能性のある事例と位置付けられる。今後は、本研究で得られた知見を基に、音楽生成モデルにおける音高・音色の制御性向上への応用や、人間の音高知覚との対応関係を検証する実験的研究へと発展させることが期待される。

謝辞： 本研究は、JST 創発的研究支援事業 JP-MJFR226V の支援を受けて実施した。

参考文献

- [1] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Sutskever, “Jukebox: A generative model for music,” *arXiv preprint arXiv:2005.00341*, 2020.
- [2] A. Agostinelli, T. I. Denk, Z. Borsos, J. Engel, M. Verzett, A. Caillon, Q. Huang, A. Jansen, A. Roberts, M. Tagliasacchi, M. Sharifi, N. Zeghidour, and C. Frank, “MusicLM: Generating music from text,” *arXiv preprint arXiv:2301.11325*, 2023.
- [3] J. Copet, F. Kreuk, I. Gat, T. Remez, D. Kant, G. Synnaeve, Y. Adi, and A. Défossez, “Simple and controllable music generation,” in *NeurIPS*, 2023.
- [4] Y. Li, R. Yuan, G. Zhang, Y. Ma, X. Chen, H. Yin, C. Xiao, C. Lin, A. Ragni, E. Benetos *et al.*, “MERT: Acoustic music understanding model with large-scale self-supervised training,” *arXiv preprint arXiv:2306.00107*, 2023.
- [5] W.-H. Liao, Y. Takida, Y. Ikemiya, Z. Zhong, C.-H. Lai, G. Fabbro, K. Shimada, K. Toyama, K. W. Cheuk, M. A. Martínez-Ramírez, S. Takahashi, S. Uhlich, T. Akama, W. Choi, Y. Koyama, and Y. Mitsufuji, “Music foundation model as generic booster for music downstream tasks,” *TMLR*, 2025. [Online]. Available: <https://openreview.net/forum?id=kHl4JzyNzF>
- [6] R. Castellon, C. Donahue, and P. Liang, “Codified audio language modeling learns useful representations for music information retrieval,” in *ISMIR*, 2021.

- [7] 八木 颯斗 and 高道 慎之介, “音楽基盤モデルは音高情報を螺旋構造に埋め込むか?” in 情報処理学会 音楽情報科学研究会, Aug. 2025.
- [8] R. N. Shepard, “Geometrical approximations to the structure of musical pitch,” *Psychological Review*, vol. 89, no. 4, pp. 305–333, 1982.
- [9] C. Limberg, F. Schulz, Z. Zhang, and S. Weinzierl, “Pitch-Conditioned Instrument Sound Synthesis From an Interactive Timbre Latent Space,” in *DAFx*, 2025.
- [10] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. Springer-Verlag, 1998.
- [11] S. Mcadams, “Perspectives on the contribution of timbre to musical structure,” *Comput. Music J.*, vol. 23, no. 3, p. 85–102, Sep. 1999. [Online]. Available: <https://doi.org/10.1162/014892699559797>
- [12] M. Wei, M. Freeman, C. Donahue, and C. Sun, “Do music generation models encode music theory?” in *ISMIR*, 2024.
- [13] V. Lostanlen, S. Sridhar, B. McFee, A. Farnsworth, and J. P. Bello, “Learning the helix topology of musical pitch,” in *IEEE ICASSP*, 2020, pp. 11–15.
- [14] S. Sridhar and V. Lostanlen, “Helicality: An isomap-based measure of octave equivalence in audio data,” in *ISMIR*, 2020.
- [15] E. Holmes, E. E. Kinghorn, L. M. McGarry, E. Busari, T. D. Griffiths, and I. S. Johnsrude, “Pitch discrimination is better for synthetic timbre than natural musical instrument timbres despite familiarity,” *The Journal of the Acoustical Society of America*, vol. 152, no. 1, pp. 31–42, 2022.
- [16] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *ACM SIGKDD*, 2019, pp. 2623–2631.

付 録

A.1 Optuna によるパラメータ探索範囲

本付録では、表 A.1 に、各データセット条件に対して用いたパラメータ探索範囲を示す。本実験では、式 (1) のパラメータの最適化に、ベイズ最適化フレームワークである Optuna [16] を用いている。各条件ごとに全楽器・調波音に対して同一の探索範囲を用いた。

最適な試行回数と繰り返し回数の検討については、我々の先行研究 [7] の付録を参照されたい。

A.2 3次元乱数の Helicity スコア

本文では、実楽器で観測された Helicity スコアが偶然的な点群配置による見かけの螺旋適合ではないことを確認するため、3次元乱数に対しても同一の手順で Helicity スコア を算出し、基準値を求めた。本付録ではその概要と、本文で用いた基準値の導出を述べる。

音高数に対応するサンプル数を $N_{\text{key}} = 36$ とし、36 個の 3次元乱数からなる点群を生成した。この点群に対して、本文と同一条件でパラメトリックな螺旋モデルをフィッティングし、Helicity スコア (式 (2)) が最大となる条件を探索した。以上の処理を独立な 3次元乱数に対して 100 試行を行い、得られた層最大 Helicity スコアの分布を集計した。

表 A.2 に、3次元乱数に対して得られた Helicity スコアの統計を示す。中央値は 0.371 であり、本文 (図 8) ではこの値を 3次元乱数に対する Helicity スコア の基準値として用いた。すなわち、実楽器で得られた Helicity スコアがこの基準値を十分に上回るかどうかを、内在音高螺旋が偶然的配置ではなく系統的に形成された構造として解釈可能かを判断する目安とした。

A.3 層別 Helicity スコア解析の補足

本付録では、本文 4.3.2 節で示した層別 Helicity スコア解析を補足するため、楽器カテゴリ別のピーク層分布および代表的楽器における層別 Helicity スコアの挙動を詳細に示す。

A.3.1 楽器カテゴリ別の層別平均 Helicity スコア

本文 4.3.2 節では、全楽器における層別 Helicity スコアの平均曲線および螺旋明瞭群におけるピーク層

分布を用いて、内在音高螺旋が主として Transformer の深い層で顕在化する傾向を示した。ここでは、その層別挙動が楽器カテゴリによってどのように異なるかを調べるため、General MIDI に基づく楽器カテゴリごとに層別 Helicity スコアの平均を算出した。

図 A.1 に示すように、多くの楽器カテゴリにおいて Helicity スコアは層の進行とともに緩やかに上昇し、第 50 層以降で顕著な増大を示す。この後半層でのピークは、内在音高螺旋が高次の表現空間において最も明瞭に形成されることを示しており、本文で示した全体傾向と整合的である。

一方で、多くの楽器カテゴリにおいて、後半層に明確なピークを持つ楽器に付随する形で、第 20 層から第 30 層付近に小さな局所的ピークが観測される。これらの浅い層におけるピークはいずれも後半層のピークと比較して僅かであり、単独で顕著な螺旋構造を形成するものではない。しかしながら、この傾向は複数の楽器カテゴリに共通して見られ、特に brass timbres および organ timbres において比較的明瞭に観測される。

カテゴリ間の差異に着目すると、bass timbres, piano timbres, guitar timbres では後半層において明確なピークが形成される一方で、pipe timbres では層方向の変化が比較的小さく、全体として低い値に留まる傾向が見られる。また、organ timbres では前半層から比較的高い値を示し、後半層にかけてさらに増大するなど、層方向に広く比較的明瞭な内在音高螺旋が分布する特徴が観測される。

これらの結果は、内在音高螺旋の顕在化が深い層で生じるという大域的傾向が楽器カテゴリを超えて一貫している一方で、浅い層における弱い構造の有無や、後半層でのピークの鋭さにはカテゴリ依存の違いが存在することを示している。

A.3.2 代表的楽器における層別 Helicity スコアの挙動

本文 4.3.2 節では、層別 Helicity スコアの平均的傾向およびピーク層の分布を用いて、内在音高螺旋が主として深い層で顕在化することを示した。しかし、個別楽器に着目すると、層方向の挙動には多様なパターンが存在する。本節では、その代表例として 4 種類の楽器を選び、層別 Helicity スコアの推移を可視化する。

図 A.2 に示すように、hammond organ では後半層を中心に高い Helicity スコアが広い範囲で維持されており、特定の単一層に鋭いピークが集中するというよりも、層全体にわたって安定した音高螺旋構

表 A.1: Optuna [16] による螺旋モデルパラメータの探索範囲。

データセット条件	θ	ϕ	h_{pitch}	r_0	r_{slope}	ω_{chroma}	t_0	h_0
SynTheory (1000 試行 × 3 回)	$[0, \pi]$	$[-\pi, \pi]$	$[0.04, 0.17]$	$[0.9, 2.1]$	$[-0.04, 0.04]$	$[-\frac{\pi}{2}, -\frac{\pi}{6}] \cup [\frac{\pi}{6}, \frac{\pi}{2}]$	$[-6.0, 6.0]$	$[-2.7, 2.0]$
SynTheory (2000 試行 × 3 回)	$[0, \pi]$	$[-\pi, \pi]$	$[0.03, 0.20]$	$[0.7, 2.5]$	$[-0.05, 0.05]$	$[-\frac{\pi}{2}, -\frac{\pi}{6}] \cup [\frac{\pi}{6}, \frac{\pi}{2}]$	$[-6.0, 6.0]$	$[-3.0, 2.0]$
調波制御音 (1000 試行 × 3 回)	$[0, \pi]$	$[-\pi, \pi]$	$[0.04, 0.17]$	$[0.9, 2.1]$	$[-0.04, 0.04]$	$[-\frac{\pi}{2}, -\frac{\pi}{6}] \cup [\frac{\pi}{6}, \frac{\pi}{2}]$	$[-6.0, 6.0]$	$[-2.7, 2.0]$

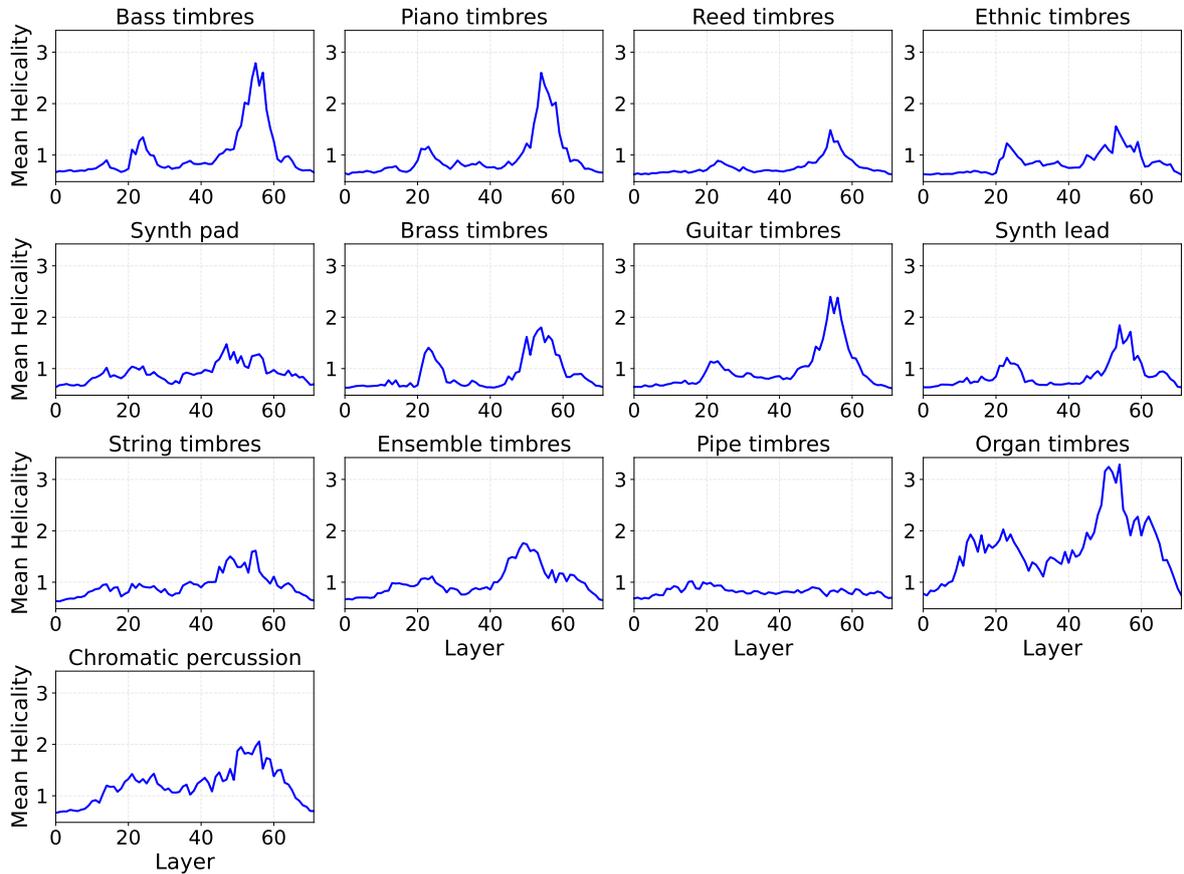


図 A.1: 楽器カテゴリごとの層別平均 Helicity スコア。各曲線は、当該カテゴリに属する楽器の Helicity スコアを層ごとに平均したものである。

表 A.2: 3次元乱数 ($N_{key} = 36$) に対する Helicity スコアの統計 (100 試行)。

統計量	Helicity スコア
平均	0.371
中央値	0.371
標準偏差	0.014
最小値 / 最大値	0.343 / 0.411

造が形成されていることが分かる。一方, slap bass 2 では第 55 層付近に鋭い最大ピークが観測され, 内在音高螺旋が特定の層で集中的に顕在化する例である。

Harpichord では, 後半層に最大ピークを持つと同時に, 第 20 層前後にも比較的高いサブピークが観測される。このような傾向は, 本文で示した層別平均曲線における前半層の局所的な上昇と対応している。

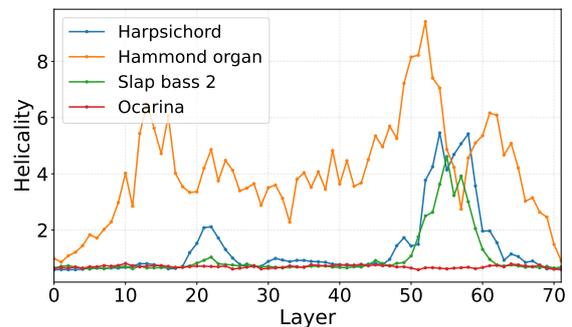


図 A.2: 代表的楽器における層別 Helicity スコアの例。

これに対し, ocarina では全層にわたって Helicity スコアが低く, 明確なピーク層を定義することが困難である。